AD-A243 853

(2)

Technical Report 1451
September 1991

Statistical Decision Making with Uncertain and Conflicting Data

R. A. Dillard

NAVAL OCEAN SYSTEMS CENTER
San Diego, California 92152-5000

J. D. FONTANA, CAPT, USN
Commander

R. T. SHEARER, Acting
Technical Director

ADMINISTRATIVE INFORMATION

This work was performed for the Independent Research Program at the Naval Ocean Systems Center, with funding provided by the Office of Chief of Naval Research (OCNR-10P). The work was carried out by the Decision Support and AI Systems Branch (Code 444) of the Naval Ocean Systems Center, San Diego, CA 92152-5000.

Released by
D. C. Eddington, Head
Decision Support and
AI Systems Branch

Under authority of
J. D. Grossman, Head
Advanced C2 Technologies
Division

PK

OBJECTIVE

Investigate automated decision-making techniques for use in situations where incoming tactical reports are frequently inaccurate or incorrect and where information about probability models is limited.

RESULTS

Of the four algorithms compared, none was satisfactory for all decision problems. Several statistical techniques were investigated for measuring the relevance of evidence to a decision problem and for identifying suspicious or conflicting evidence. An approach was described for two representative problems. A simple version was implemented in CLIPS to uncover unforeseen difficulties that could occur. The approach seems feasible, although it depends on having current probability assignments or other numerical judgments from experts. While maintaining current probability distributions agreed upon by experts may be the most difficult part of the problem, it should be feasible with the help of a user-friendly, interactive program.

CONTENTS

FIGURES

TABLES

1.0 BACKGROUND

Basing a decision on data from unreliable sources is a common experience. Bayesian arguments (Pearl, 1988; Ng & Abramson, 1990) can provide an optimum solution but require prior and conditional probabilities that are often unknown. We are investigating automated decision-making techniques suitable for situations where incoming reports are frequently inaccurate or incorrect and where information about probability models is limited.

1.1 THE DECISION PROBLEM

The basic problem is to decide among an exhaustive set $\{Hi\}$ of mutually exclusive hypotheses, given a set of evidence $\{Ej\}$. The evidence consists of reported facts or measurements and, for each, we have a measure Wj of its accuracy or confidence in its source, as discussed in the next section. Numerical judgments of $\{Hi\}$ can have any of four forms. For hypothesis Hi and evidence Ej, we may have $P(Hi|Ej)$, $P(Ej|Hi)$, a fuzzy measure of how consistent evidence Ej is with Hi, or a Dempster-Shafer probability assignment $mj(Hi)$ (Dempster, 1967; Shafer, 1976). We generally will not have estimates of other probabilities needed to perform Bayesian analysis, e.g., the probability of Hi conditional on Ej not occurring, $P(Hi|\neg Ej)$.

The candidate decision methods compared here are the linear opinion pool, the logarithmic opinion pool, Dempster's rule (Dempster, 1967), and a fuzzy-logic algorithm. Of these, only Dempster's rule requires independent evidence. Comparisons include the MYCIN (Shortliffe, 1976) certainty-factor calculus for the two-hypothesis examples, where $H2$ is $\neg H1$ and the numerical judgments are measures of belief and disbelief in $H1$.

1.2 DATA-QUALITY MEASURES

The kinds of data pertinent to tactical decision problems include parameter measurements and statements of fact. Some of the facts are derived from raw parameter measurements, such as facts concerning emitter classifications. Reasons for bad tactical data include human error, faulty equipment, sensor resolution, navigation error, contact ambiguity, and midlevel data-fusion errors.

The definition of a measure of data quality will depend on data type and on how that measure is to be used. The study described in Section 2 is to determine appropriate ways of using the measures. Separate studies are directed toward defining and measuring tactical data quality. The comparison of decision methods is valid for the values of Wj arbitrarily specified, and calibrating the measure to the decision problem can be treated as a separate issue. However, we need to at least consider here the nature of a data-quality measure.

The simplest definition of a data-quality measure for factual data is the probability that the fact is true, i.e., $Wj = P(Ej \text{ true})$. However, unless the measure is well over 0.5, we also would need to know $P(Hi|\neg Ej)$. In practice, a small measure often will imply uncertainty, ignorance, or ambiguity more than incorrectness. For example, an observer may be uncertain about the class of a ship seen in the dark or in a heavy fog, or about the classification of a signal embedded in noise. Also, an intercepted signal may be used to classify a contact, but the signal may actually have been emitted by another vessel in the same general direction. This is especially likely to occur when navigation errors are large or when atmospheric ducting conditions exist and the emitting vessel is well beyond the horizon. A knowledge-based system could estimate the data quality based on the spatial density of other contacts and on atmospheric conditions in addition to sensor performance. The definition of data quality also depends on use. For example, gross inaccuracies in position may be fully acceptable for determining that a contact is located in a commercial lane, a common path of transit, or in an unusual area, because the position error is usually small compared to lane width or area size.

1.3 NUMERICAL PROBABILITY JUDGMENTS

The user of a decision algorithm must supply numerical judgments of the hypotheses. Some of the evidence contributes more naturally to probability distributions of the form $\{P(Hi|Ej)\}$ or $\{mj(Hi)\}$, discussed later in this section, and other evidence to $\{P(Ej|Hi)\}$ or to fuzzy measures of expectation of Ej for each hypothesis Hi. To compare a fuzzy method with the others, we have interpreted the conditional probabilities $\{P(Ej|Hi)\}$ to be a fuzzy subset of the set $\{Hi\}$, where $P(Ej|Hi)$ indicates to what degree evidence Ej is consistent with Hi. This is similar to the way these conditional probabilities would be given values by a domain expert.

Conversion of the distribution $\{P(Ej|Hi)\}$ to $\{P(Hi|Ej)\}$ for evidence Ej is straightforward, provided that care is given to the interpretation of the resulting probabilities and to the possible interplay of prior or base-rate probabilities. The relationship is found by using Bayes' rule:

$$P(Hi|Ej) = P(Ej|Hi)\frac{P(Hi)}{\sum_k P(Ej|Hk)P(Hk)} \quad . \tag{1}$$

When we can treat the prior probabilities $\{P(Hi)\}$ as equal, we have

$$P(Hi|Ej) = \frac{P(Ej|Hi)}{\sum_k P(Ej|Hk)} \quad . \tag{2}$$

Note that in Equation 2, the probabilities $\{P(Hi|Ej)\}$ cannot be converted to $\{P(Ej|Hi)\}$ without knowledge of $P(Ej|Hi)$ for at least one Hi.

There are constraints on the collective probability distributions and data-quality measures. Assume that the measures $\{Wj\}$ are on a scale of 0 to 1, where $Wj = 1$ means that the information from the jth source is absolutely correct. (We do not allow $Wj = 0$.) Recall that the hypotheses are exhaustive and mutually exclusive, giving $\sum_i P(Hi|Ej) = 1$. We disallow the following combination, because one piece of evidence Ej rules out the hypothesis Hi and the other, El, rules out all other hypotheses:

$$[Wj = 1;\ P(Hi|Ej) = 0] \text{ and } [Wl = 1,\ l \neq j;\ P(Hi|El) = 1].$$

Another combination we disallow, because it implies the above, is

$$[Wj = 1;\ P(Hk|Ej) = 1] \text{ and } [Wl = 1,\ l \neq j;\ P(Hi|El) = 1,\ k \neq i].$$

In all of our applications, what could be considered a prior distribution is treated as a base-rate frequency (Tverksy & Kahneman, 1974), which is a probability distribution contributed by only one piece of evidence. A base rate for a radar or sonar contact typically results from its location in a particular area or commercial lane. For example, suppose that the hypotheses correspond to the ship type of a radar contact and that evidence $E1$ puts the contact in a particular merchant lane. If $H15$ corresponds to Merchant, then $P(H15|E1)$ is the ratio of the long-term count of merchants in that lane to the long-term count of all platforms in the lane, and similarly for other ship types. The conditional probabilities contributed by all other evidence should not be affected by this base rate, so we can use Equation 2 to convert $P(Ej|Hi)$ to $P(Hi|Ej)$.

Many of the conditional distributions contributed by evidence will be derived subjectively or will be based on recollection of the frequency of outcomes in similar cases. Some can be derived empirically by using probability distributions of sensory measurements (Garvey et al., 1981). In a later example, representative curves are derived empirically from sensor measurements of initial detection range and speed for each ship type.

To compare Dempster's rule with the other methods, we allow probability to be assigned only to the singleton hypotheses and to the disjunction U of all hypotheses, i.e., to $H1$ OR $H2$ OR $H3$ .... (In the general application of Dempster's rule, probability can be assigned to any disjunction of hypotheses.) The probability mass assignment (that we call it to distinguish it from the other distributions) based on evidence Ej then consists of $\{mj(Hi)\}$ and $mj(U)$. We can simply let $mj(Hi) = 1 - Wj$, where the Wj's are on a scale of 0 to 1, and let $mj(Hi)\} = Wj \bullet P(Hi|Ej)$; however, we would overlook a primary reason for using Dempster's method. The probability $mj(U)$ can represent uncertainty concerning both the evidence and the interpretation of it, and representing the latter kind of uncertainty is also important. In most of the examples of applications that we give later, the original probability distributions consist of $\{mj(Hi)\}$ and $mj(U)$, where $mj(U)$ represents both kinds of uncertainty. We later address the problem of converting these to conditional probabilities for use in the other algorithms.

1.4 ISSUES

Section 2 addresses the need for a decision algorithm that uses numerical probability judgments and data-quality measures. We could find no decision algorithm intended for this purpose, but several appear to be applicable, with some reinterpretation. A question is if an existing algorithm will perform adequately or must one be designed for this purpose?

Another question is how to determine which data to use in a decision problem or, equivalently, which decision problems need each datum. Section 3 discusses ways of measuring the relevance of evidence to a set of hypotheses, and also ways of measuring conflict among the pieces of evidence with respect to a set of hypotheses. Data-quality measures simply indicate the average quality of the source, after adjustments for propagation and other factors. The actual quality may fluctuate widely for some kinds of data. Statistical techniques can be used to identify conflicting or suspicious data needing verification.

Finally, is it realistic to expect that these statistical techniques and other needed inferencing can be implemented in a knowledge-based system? Section 4 investigates possible system architectures and discusses how to obtain the needed numerical judgments.

2.0 USING DATA-QUALITY MEASURES IN DECISION-MAKING ALGORITHMS

In this section, we will investigate the performance of several algorithms in which we can use data-quality measures. While none of these methods was intended for quite this purpose, all are applicable with certain limitations.

2.1 DECISION METHODS

For each of the decision methods below, select hypothesis *Hi* such that the decision statistic *di* is greatest (or rank the hypotheses in the order of the values of *di*). In the first three methods, normalize the weights *Wj* to obtain weights *wj* such that $\sum wj = 1$. The two opinion-pool methods are described in Genest and Zidek (1986) and in Zidek (1986). The weighted Min-Max is a fuzzy logic method that Yager (1977) introduced for another kind of application, where *Wj* represented criterion importance. Dempster's rule (Dempster, 1967; Shafer, 1976) requires independent evidence.

2.1.1 Linear Opinion Pool

$$di = \sum_j wj \bullet P(Hi|Ej) \tag{3}$$

2.1.2 Logarithmic Opinion Pool

$$di = \frac{\prod_j P(Hi|Ej)^{wj}}{\sum_i \prod_j P(Hi|Ej)^{wj}} \tag{4}$$

A small value of *wj* results in a factor close to 1 for all *Hi*, which reduces the effect of that evidence on the decision statistic *di*. Figure 1 gives exponentially weighted probabilities for a variety of values of exponent *w*. The weights are traditionally normalized to sum to 1, but normalization does not affect the ordering of the values of *di*. For normalized weights, note that if all the evidence gives the same conditional distribution $\{P(Hi|E)\}$, then $di = P(Hi|E)$. This is also true for the linear opinion pool, and is a desirable property when pooling opinions that might be based on the same evidence.

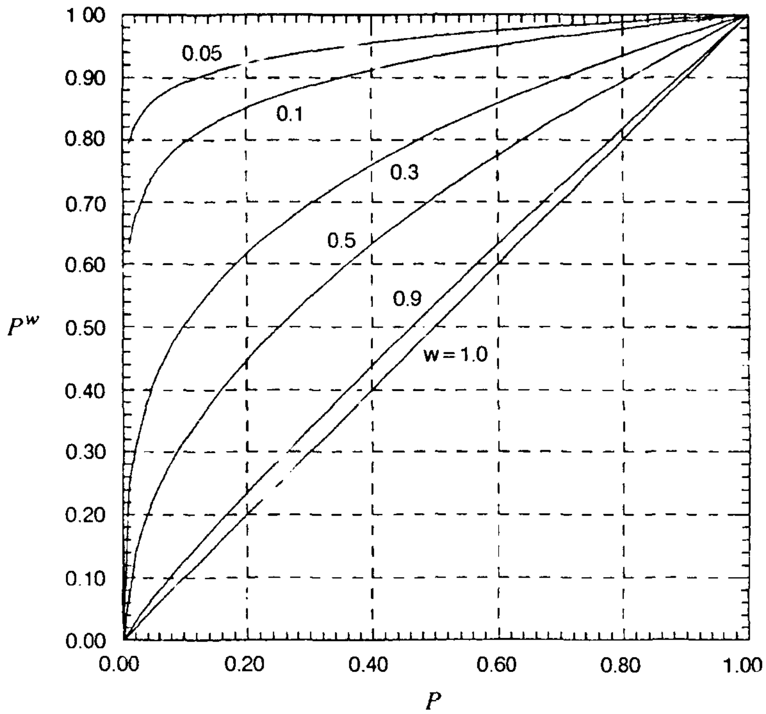

Figure 1. Effect of exponential weighting.

2.1.3 Weighted Min-Max

$$di = \min_j cji^{wj} \tag{5}$$

The fuzzy measure cji is the degree to which evidence Ej is consistent with hypothesis Hi. ($Cj = \{cji\}$ is a fuzzy subset of $H = \{Hi\}$.) For comparisons with other methods, let $cji = P(Ej|Hi)$. If we assume equal prior probabilities, the decision ordering is the same whether we use $P(Ej|Hi)$ or $P(Hi|Ej)$. Figure 1 applies also to this method. Note that a very small value of wj is unlikely to produce a minimum value of cji^{wj}.

2.1.4 Dempster's Rule

As stated earlier, the probability mass assignment based on evidence Ej consists of $\{mj(Hi)\}$ and $mj(U)$, where U is the disjunction of all hypotheses. The lower bound to the probability $P(Hi|\{Ej\})$ that Hi is true based on all evidence is (Dillard, 1982)

$$m(Hi) = \left\{ \prod_j [mj(Hi) + mj(U)] - \prod_j mj(U) \right\} / C = F(Hi)/C \tag{6}$$

where $F(Hi)$ represents the expression in braces and

$$C = \prod_j mj(U) + \sum_i F(Hi) . \tag{7}$$

The resulting uncertainty is

$$m(U) = \prod_{j} mj(U)/C \tag{8}$$

The resulting upper bound to $P(Hi|\{Ej\})$, which is known as the plausibility of Hi, is

$$p(Hi) = m(Hi) + m(U) \ . \tag{9}$$

Let the decision statistic be

$$di = m(Hi) + m(U)/n \ , \tag{10}$$

where n is the number of hypotheses. The values of di then sum to 1. (This statistic does not apply to the general case where probability mass is assigned to disjunctions in addition to U.)

Conversion from mass assignment to conditional probabilities. The probability assigned to uncertainty U represents uncertainty concerning both the evidence and the expert's interpretation of it. The other decision methods do not allow the expert to express uncertainty about the conditional probabilities. We could change this by redefining the weight Wj to include both kinds of uncertainty, in which case the conversion formulas for comparison purposes are $Wj = 1 - mj(U)$ and $P(Hi|Ej) = mj(Hi)/Wj$, where the $\{Wj\}$ are specified on a scale of 0 to 1. A serious problem with these formulas can occur, though, when $mj(Hi) = 0$, as discussed in Section 2.1.5. To avoid this problem and to maintain our concept of Wj as a data-quality weight, we give here a conversion method that creates Wj from only that component of $m(U)$ that represents data quality.

Separate $mj(U)$ into two components, $mj(U) = mj_E(U) + mj_H(U)$, where $mj_E(U)$ represents uncertainty about the truth or accuracy of the evidence and $mj_H(U)$ represents uncertainty about how to interpret the evidence, assuming it is accurate or true. Let the data-quality measure for evidence Ej be

$$Wj = 1 - mj_E(U) \tag{11}$$

and the probability of Hi conditional on Ej be

$$P(Hi|Ej) = \frac{mj(Hi) + mj_H(U)/n}{1 - mj_E(U)} \ , \tag{12}$$

where n is the number of hypotheses.

Conversion to mass assignment from conditional probabilities. Let $mj_E(U) = 1 - Wj$, where the data quality weights Wj are specified on a scale of 0 to 1. Optionally, assign a value to $mj_H(U)$ to represent uncertainty about the interpretation of evidence Ej, i.e., uncertainty about the probability assignment $\{P(Hi|Ej)\}$. The default value is 0. Let

$$mj(U) = mj_E(U) + mj_H(U) \tag{13}$$

and

$$mj(Hi) = Wj \bullet P(Hi|Ej) - mj_H(U)/n \; . \tag{14}$$

2.1.5 Desired Properties

Assume that the $\{Wj\}$ are scaled from 0 to 1 and that $Wj = 1$ implies that Ej is absolutely true. We desire the following properties in a decision algorithm.

Property 1. $Wj = 1$; $P(Hi|Ej) = 0 => Hi$ rejected.

Property 2. $Wj = 1$; $P(Hi|Ej) = 1 => Hi$ selected (because $P(Hk|Ej) = 0$ for all $k \neq$ i).

Property 3. $Wj < 1$; $P(Hi|Ej) = 0 => Hi$ not automatically rejected.

Property 4. $P(Hi|Ej) > P(Hk|Ej)$ for all $k \neq$ i and all $j => Hi$ selected.

Table 1 shows which properties are satisfied by the four decision methods. Properties 1 and 2 are desirable only if we assume that the domain expert will not give $P(Hi|Ej)$ a value of 0 or 1 in error. The expert may be uncertain of intermediate values but knows when the evidence, if correct, rules out a hypothesis or can occur only with one hypothesis. The results for Dempster's rule assume that the conversion is made by letting $m(U) = 1 - Wj$ and $mj(Hi) = Wj \bullet P(Hi|Ej)$; that is, the component $mj_H(U)$ is zero.

Table 1. Properties satisfied or not satisfied by decision algorithms.

Desired Property	Linear Opinion	Log Opinion	Dempster's Rule	Weighted Min-Max Pool
1	No	Yes	Yes	Yes
2	No	No	Yes	Yes
3	Yes	No	Yes	No
4	Yes	Yes	Yes	Yes

By not satisfying Property 3, the logarithmic opinion pool and the fuzzy logic method give veto power to an individual piece of evidence. Note that if all of the hypotheses are vetoed, di is not defined in the logarithmic pool and is zero for all i in the fuzzy method. Very small values of $P(Hi|Ej)$ or $P(Ej|Hi)$ have an effect close to vetoing. Because it is based on a product, the logarithmic pool is particularly sensitive to errors in estimating small conditional probabilities, e.g., estimating 0.005 when 0.01

is appropriate. The logarithmic pool fails Property 2 only because some other evidence Ek with $Wk < 1$ can contribute $P(Hi|Ek) = 0$ and veto Hi.

With Dempster-Shafer mass assignments, it is common that $mj(Hi)$ is nonzero when evidence Ej increases belief in Hi and is zero when Ej does not. Because letting $E(Hi|Ej)$ be $mj(Hi)/Wj$ yields vetoes for all Hi such that $mj(Hi) = 0$ for some j, that conversion is impractical for use with the logarithmic pool and the fuzzy method. In fact, in the two-hypothesis examples given later, every hypothesis would be vetoed. For this reason, we use Equation 14 in those examples.

A desirable property when the evidence is mutually independent is that of reinforcement. With Dempster's rule, for example, if all evidence contributes the same probability assignment, say with the greatest probability assigned to $H1$ and the least to $H2$, then $d1$ will be significantly greater than $mj(H1) + mj(U)/n$ and $d2$ will be significantly less than $mj(H2) + mj(U)/n$. Recall that the opinion pools would give $d1 = P(H1|Ej)$ and $d2 = P(H2|Ej)$.

2.1.6 Certainty Factors

The certainty-factors calculus used in the MYCIN (Shortliffe, 1976) system is applicable to our two-hypothesis decision problems, where $H2$ is $\neg H1$ (because the hypotheses are exhaustive and mutually exclusive). A certainty factor is a number between 1 and -1, where 1 means $H1$ is absolutely true and -1 means $H1$ is absolutely false, i.e., that $H2$ is true. The certainty factor for evidence Ej is defined by $CF[H1,Ej] = MB[H1,Ej] - MD[H1,Ej]$, where $MB[H1,Ej]$ is the measure of increased belief in $H1$ based on the evidence Ej and $MD[H1,Ej]$ is the measure of increased disbelief in $H1$ based on Ej. The combined certainty factor is $CF = d1 - d2$, where $d1$ and $d2$ are the resulting measures of belief in $H1$ and $H2$, respectively, and are computed as shown below. Let m be the number of pieces of evidence, and assume that no measure of belief or disbelief is give a value of 1.

$$
\begin{aligned}
d_{1,1} &= MB[H1,E1] \\
d_{1,k} &= d_{1,k-1} + MB[H1,Ek](1 - d_{1,k-1}) \\
d_1 &= d_{1,m} \\
d_{2,1} &= MD[H1,E1] \\
d_{2,k} &= d_{2,k-1} + MD[H1,Ek](1 - d_{2,k-1}) \\
d_2 &= d_{2,m}
\end{aligned}
$$

Rather than attempt to define a relationship between certainty factor assignments and Dempster-Shafer probability mass assignments for the purpose of comparisons, we simply let $MB[H1,Ej] = mj(H1)$ and $MD[H1,Ej] = mj(H2)$. In every assignment in the two-hypothesis examples, one of these measures is zero, which is a requirement in the certainty factor method. Uncertainty about the evidence and its interpretation is assumed embodied in the belief measures.

2.2 DECISION PROBLEM EXAMPLES

2.2.1 Iranian Airbus Incident

The first example is based on the Iranian Airbus incident as reported in Friedman (1989), Carlson (1989), and Crowe (1989). A commercial aircraft, departing the Bandar Abbas airfield on the morning of 3 July 1988, was identified by USS *Vincennes* personnel as a hostile F-14 and was destroyed by a missile. Data contributing to the identification decision were (1) the aircraft originated from an airfield used both by commercial and military aircraft but was off the center line of the commercial air corridor by 3 or 4 miles, (2) the aircraft's altitude was reported to be declining as it approached the *Vincennes*, and (3) a Mode II IFF squawk (associated with an Iranian F-14) was detected and attributed to the aircraft.

Table 2 gives arbitrary values of data-quality weights Wj and conditional probabilities $P(Ej|Hi)$ for four hypotheses. The report that the aircraft was descending was in error so is given a low weight (probably much lower than its credibility at the time). Values of $W2$ and $P(E2|Hi)$ are also given for a what-if case, where altitude is correctly reported. We treat as separate evidence two aspects of the IFF squawk. First, only certain aircraft have the capability of emitting Mode II. Some Iranian commercial aircraft were believed to have that capability because they were also used militarily. Secondly, is there a good reason why, under each hypothesis, the contact would have its IFF energized? (The IFF squawk was not from the airbus, and the fact that $P(E4|Hi)$ is small for all Hi could be used as an indicator of bad data.)

Facts not taken into account here include earlier military incidents. These incidents significantly increased the expectation of a hostile event, which probably was normally quite low. Another fact not specifically taken into account was that the airbus takeoff was 27 minutes behind any scheduled commercial departure. While this apparently contributed to the decision to react, such a delay seems normal in our experience. However, we could consider this fact to be included in evidence $E1$.

Table 2. Probabilities (fictitious) for the Irania ı airbus incident.

	Wj and $P(Ej\|Hi)$				
Evidence Ej:	$E1$: Origin and location (off center) in corridor	$E2$: De- scending	[Not De- scending]	$E3$: Mode II IFF (capability of)	$E4$: Mode II IFF (reason for emitting)
Quality Wj:	0.9	0.2	[0.8]	0.3	0.3
Hypothesis					
$H1$: Military–Hostile intent	0.8	0.7	[0.3]	0.7	0.1
$H2$: Military–Innocent passage	0.4	0.2	[0.8]	0.9	0.4
$H3$: Commercial carrier	0.7	0.05	[0.95]	0.1	0.05
$H4$: Private or other commercial	0.6	0.2	[0.8]	0.7	0.05

Conversion of $P(Ej|Hi)$ into $P(Hi|Ej)$ for the opinion pools used Equation 2. Further conversion into probability mass assignments for Dempster's rule used $mj(U) = 1 - Wj$ and $mj(Hi) = Wj \bullet P(Hi|Ej)$.

Table 3 gives results for the case where evidence $E2$ is the false report that the aircraft was descending. The decision order for Dempster's rule and the linear pool is the same: $H1$, $H2$, $H3$, $H4$. The logarithmic pool has a reverse order for $H3$ and $H4$, while the fuzzy method has a tie between $H3$ and $H4$. The values of di for the two opinion pools fall within Dempster's lower and upper bounds (an uncertainty range of 0.072) only for the first choice, $H1$. For the other hypotheses, some fall above and some below.

Table 3. Comparison of decisions for Iranian airbus incident.

	di and Decision order				
Hypothesis	Dempster's rule	[with prob. bounds]	Linear pool	Log pool	Fuzzy
$H1$ Military -	0.359	[0.341–0.413]	0.351	0.386	0.666
Hostile intent	1		1	1	1
$H2$ Military -	0.228	[0.210–0.282]	0.303	0.295	0.616
Innocent passage	2		2	2	2
$H3$ Commercial	0.217	[0.199–0.271]	0.177	0.158	0.589
carrier	3		3	4	3/4
$H4$ Private or	0.196	[0.178–0.250]	0.168	0.161	0.589
Other commercial	4		4	3	3/4

Table 4 gives the results for the modified version of the incident, where the altitude was not mistakenly reported as decreasing. The ordering of the decisions for the two opinion pools is the same, and the fuzzy method shares *H*2 as the first choice. The ordering is different for Dempster's rule, but note that the values of *di* vary little among the hypotheses for all methods. In fact, the difference is less than 0.001 in two instances. Again, the values of *di* for the opinion pools fall within Dempster's bounds (an uncertainty range of 0.36) only for *H*1.

Table 4. Comparison of decisions for Iranian airbus incident, except altitude correctly reported.

Hypothesis	*di* and Decision order: Dempster's rule [with prob. bounds]		Linear pool	Log pool	Fuzzy
*H*1 Military – Hostile intent	0.2274 3	[0.218-0.254]	0.2432 2	0.245 2	0.658 4
*H*2 Military – Innocent passage	0.260 2	[0.251-0.287]	0.307 1	0.315 1	0.699 1
*H*3 Commercial carrier	0.286 1	[0.277-0.313]	0.2430 3	0.238 3	0.677 2/3
*H*4 Private or Other commercial	0.2270 4	[0.218-0.254]	0.207 4	0.202 4	0.677 2/3

2.2.2 Ship Type

The hypotheses in this example correspond to platform type, with types of lesser importance combined to keep the number of hypotheses manageably small. The hypotheses are as follows (Dillard, 1982).

*H*1: carrier
*H*2: cruiser
*H*3: destroyer
*H*4: frigate
*H*5: amphibious
*H*6: submarine (surfaced or periscope/snorkel/antenna)
*H*7: small fighting ship
*H*8: fast attack/patrol craft
*H*9: patrol craft (not fast)
*H*10: intelligence collector
*H*11: survey/research (navy operated)
*H*12: fleet auxiliary – medium & large
*H*13: fleet auxiliary – small
*H*14: small boats (navy and commercial)
*H*15: merchant

$H16$: fishing
$H17$: other commercial and private
$H18$: debris

We assume that there is one and only one platform under consideration; e.g., the radar blip is not just radar noise or clutter and is not the return from two platforms close together. We disregard the possibility that a contact that appears to be a surface platform might be an aircraft.

Table 5 gives probability mass assignments taken from Dillard (1982). Evidence $E1$ is the location of the radar contact. Base-rate frequencies are found from the long-term average numbers of each type of platform at that general location, assuming that no current military or weather event causes a change. The mass assignment for location evidence in table 5 represents these base-rate frequencies and an uncertainty mass $m1(U) = 0.35$.

Table 5. Probability mass assignments for ship type.

Hyp.	Ship type	Location $m1$(hyp)	Range $m2$(hyp)	Speed $m3$(hyp)
$H1$	carrier	0.003	0	0.103
$H2$	cruiser	0.013	0	0.103
$H3$	destroyer	0.040	0	0.103
$H4$	frigate	0.030	0	0.029
$H5$	amphibious	0.040	0	0.008
$H6$	submarine	0.020	0.037	0.009
$H7$	small fighting	0.040	0.045	0.029
$H8$	fast a/p craft	0.060	0.136	0.074
$H9$	patrol craft	0.020	0.136	0.009
$H10$	intel. collector	0.015	0.027	0.008
$H11$	survey/research	0.012	0	0.008
$H12$	auxiliary-med/lrg	0.090	0	0.008
$H13$	auxiliary-sml	0.030	0.045	0.008
$H14$	small boats	0.010	0.055	0.011
$H15$	merchant	0.130	0.023	0.034
$H16$	fishing	0.070	0.077	0.011
$H17$	other com/private	0.020	0.064	0.045
$H18$	debris	0.007	0.055	0
U	uncertainty	0.350	0.300	0.400

The initial detection range of the radar contact is 11 nautical miles, and the measurement of its speed is 31 knots. Probability mass assignments are found for these measurements by using a method similar to that used with emitter parameter distributions by Garvey, et al. (1981). Figure 2 illustrates the method with distribution

curves for range at initial detection. (Such curves are valid for only one antenna height, frequency, and environment state, and assume a low signal strength.) For each measurement and each ship type, a strip corresponding to the measurement ± the average measurement error is overlaid on the appropriate distribution curve and the overlapped area is computed. These areas are normalized to sum to the complement of the probability mass of the uncertainty, $mj(U)$, concerning the measurement and its interpretation. Letting $m2(U)$ be 0.3 for the range measurement and $m3(U)$ be 0.4 for the speed measurement gives the values in the $m2$ and $m3$ columns of table 5.

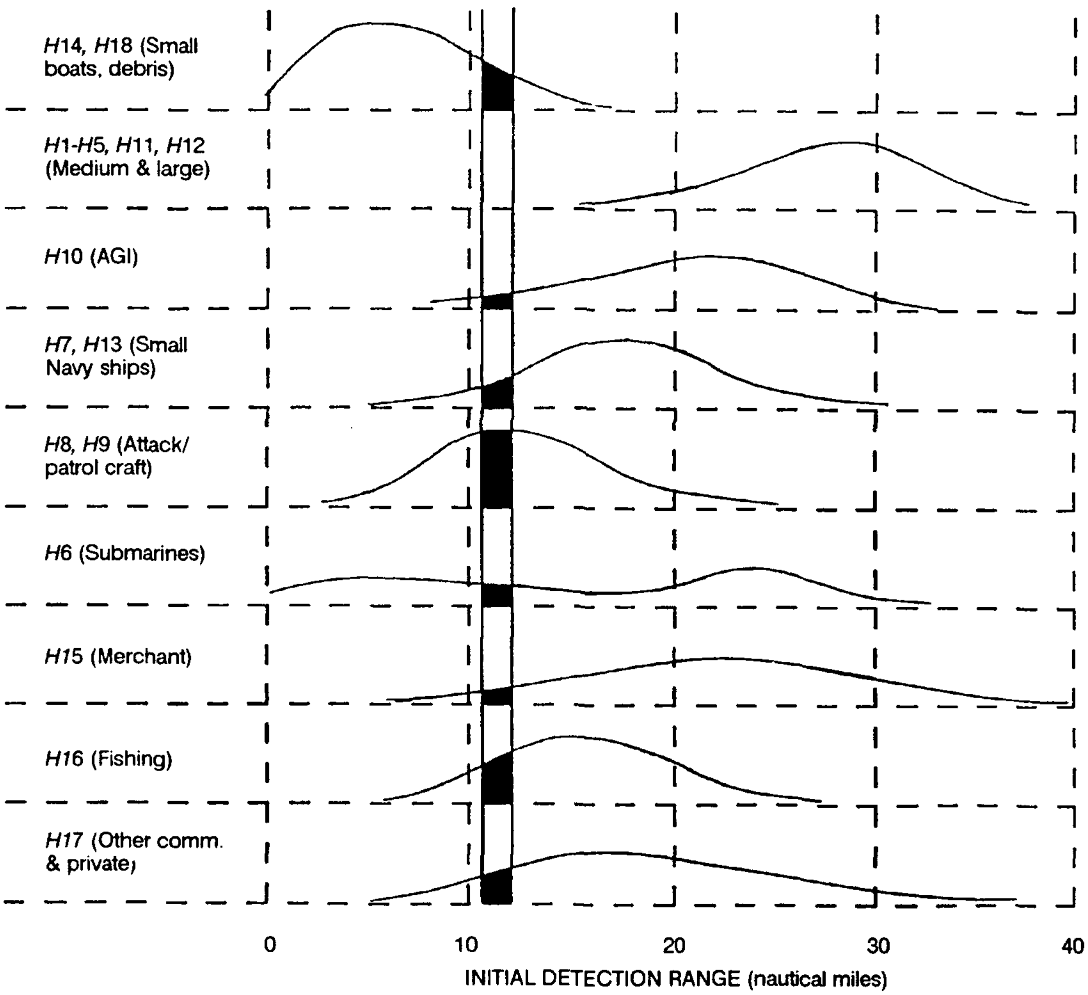

Figure 2. Distribution functions for initial detection range (Dillard, 1982).

To compare the other decision methods with Dempster's rule, we convert the three probability mass assignments of the form $mj(Hi)$ in table 5 to the conditional probabilities $P(Hi|Ej)$ by using $Wj = 1 - mj(U)$ and $P(Hi|Ej) = mj(Hi)/Wj$. That is, we let $mj_H(U)$ be 0 in equations 11 and 12. We cannot obtain $P(Ej|Hi)$ for use in the fuzzy method, but substituting $P(Hi|Ej)$ gives the same decision order. The values of di for the four decision methods are given in table 6. Note that all methods result in the same preferred decision, $H8$: *fast attack/patrol craft*. However, the values of $d8$ for Dempster's rule and the opinion pools are so small that making a decision is inadvisable, so the useful result is an ordered list of most likely types. $H15$ is the second choice for all but the fuzzy method, where it is fourth. In general, the results for all methods are fairly similar. It is interesting to note that the nonzero values of di for the opinion pools are within Dempster's bounds (an uncertainty range of 0.138) for all except $H8$, where the linear pool's $d8$ is lower. Note also that the last-place decision for all is $H11$, although it is a tie at zero with other hypotheses for the logarithmic pool and the fuzzy method.

As discussed earlier, a property of the logarithmic pool and the fuzzy method that is undesirable (unless evidence Ej is absolutely correct) is that di will be zero if the conditional probability of Hi is zero for a single piece of evidence. In this example, however, values of zero for di are not unreasonable. Debris and large ships can be ruled out even if the measurement errors are large, because debris travels much slower than 31 knots and a large ship can be detected by radar at a range much greater than 11 nautical miles under the assumed conditions.

Comparisons (not shown) were also made for the case where some of the uncertainty mass is attributed to uncertainty about the interpretation of the evidence. Equations 11 and 12 were used with $m1_H(U) = 0.25$, $m2_H(U) = 0.15$, and $m3_H(U) = 0.2$. With this conversion, no conditional probability was zero and therefore no value of di was zero. Compared to the results in table 6, the ordering of the decisions for Dempster's rule and the opinion pools was unchanged for the four highest, respectively. The decisions for the fuzzy method differed starting with the third. Overall, the results were nearly the same as in table 6.

Table 6. Values of *di* for decision on ship type.

Hi	Ship type	Dempster's rule [with prob. bounds]		Linear pool	Log pool	Fuzzy
*H*1	carrier	0.045	[0.037–0.175]	0.054	0	0
*H*2	cruiser	0.050	[0.042–0.180]	0.059	0	0
*H*3	destroyer	0.063[6]	[0.055–0.193]	0.073[5]	0	0
*H*4	frigate	0.030	[0.023–0.160]	0.030	0	0
*H*5	amphibious	0.026	[0.019–0.157]	0.025	0	0
*H*6	submarine	0.037	[0.030–0.167]	0.034	0.055	0.275*
*H*7	small fighting	0.059	[0.052–0.189]	0.058	0.106[5]	0.373[2]
*H*8	fast a/p craft	0.148[1]	[0.140–0.278]	0.138[1]	0.240[1]	0.452[1]
*H*9	patrol craft	0.086[3]	[0.079–0.216]	0.085[3]	0.087[6]	0.275*
*H*10	intel. collector	0.030	[0.022–0.160]	0.026	0.043	0.265
*H*11	survey/research	0.015[L]	[0.008–0.145]	0.010[L]	0	0
*H*12	auxiliary	0.047	[0.039–0.177]	0.050	0	0
*H*13	auxiliary-sml	0.045	[0.038–0.175]	0.043	0.065	0.265
*H*14	small boats	0.042	[0.035–0.172]	0.039	0.053	0.249
*H*15	merchant	0.091[2]	[0.083–0.221]	0.096[2]	0.129[2]	0.293[4]
*H*16	fishing	0.083[4]	[0.076–0.213]	0.081[4]	0.115[3]	0.292[5]
*H*17	other com/private	0.066[5]	[0.059–0.197]	0.066[6]	0.109[4]	0.313[3]
*H*18	debris	0.036	[0.029–0.166]	0.032	0	0

[L] - Last *tied - 6 & 7

2.2.3 Merchant Ships

Merchant ships have certain behavior characteristics. They travel in merchant lanes, rarely speeding or changing speed and rarely changing course except to avoid a storm. By contrast, military ships frequently maneuver and, during hostilities, are more likely to be within weapons range of a high-value target. Characteristics such as these are useful in deciding whether or not a radar contact is a merchant, although, of course, a military ship may purposely try to appear like one. Table 7 summarizes four situations, each with six pieces of evidence and under the assumption of open sea and clear weather. The contact is a merchant in Cases 1 and 2 and is a cruiser in Cases 3 and 4. The probabilities in table 7 are given as Dempster-Shafer mass assignments. Each assigns probability mass only to one of the two hypotheses, so this mass can also be used as a measure of increased belief or disbelief in *H*1 to compute MYCIN certainty factors (Shortliffe, 1976). (Empirically derived assignments would generally give non-zero probability mass to both hypotheses.) To convert the assignments to *Wj* and $\{P(Hi|Ej)\}$, values of both components of $mj(U)$ are specified.

Table 7. Evidence and probability masses for decision between $H1$: Merchant and $H2$: Not merchant.

Evidence Ej	$E1$: Merchant Lane	$E2$: Detection Range	$E3$: Speed	$E4$: Course Change	$E5$: Speed Change	$E6$: Within Reach
Case 1 (Typical merchant)	Inside	21 NM	22 knots	No	No	No
$mj(H1)$	.3	.15	.2	.15	.1	.3
$mj(H2)$	0	0	0	0	0	0
$mj(U)=mj_E(U) +mj_H(U)$	.7=.1+.6	.85=.3+.55	.8=.4+.4	.85=.3+.55	.9=.3+.6	.7=.5+.2
Case 2 (Large fast merchant)	Inside	28 NM	28 knots	No	No	No
$mj(H1)$	.3	0	0	.15	.1	.3
$mj(H2)$	0	.15	.1	0	0	0
$mj(U)=mj_E(U) +mj_H(U)$	.7=.1+.6	.85=.3+.55	.9=.4+.5	.85=.3+.55	.9=.3+.6	.7=,5+.2
Case 3 (Cruiser)	Inside	30 NM	30 knots	No	No	No
$mj(H1)$	.3	0	0	.15	.1	0
$mj(H2)$	0	.4	.25	0	0	.2
$mj(U)=mj_E(U) +mj_H(U)$	.7=.1+.6	.6=.3+.3	.75=.4+.35	.85=.3+.55	.9=.3+.6	.8=.1+.7
Case 4 (Cruiser)	Outside	30 NM	30 knots	Yes (6^0)	No	Yes
$mj(H1)$	0	0	0	0	.1	0
$mj(H2)$	.6	.4	.25	.5	0	.2
$mj(U)=mj_E(U) +mj_H(U)$	.4=.1+.3	.6=.3+.3	.75=.4+.35	.5=.3+.2	.9=.3+.6	.8=.1+.7

Table 8 gives the results for the four candidate methods and for the certainty factor method. The value of di for the preferred hypothesis in each case is underlined in table 8. Note that all methods decide alike in all four cases. In Case 1, Dempster's rule and the MYCIN method show reinforcement, a desirable property if the evidence is independent. The opinion pool values are outside of Dempster's bounds even though the uncertainty range is 0.255, which is fairly large. (The resulting uncertainty mass is the product of the individual masses when each distribution assigns mass to uncertainty and only one of the two hypotheses.) In Case 2, the opinion pool and MYCIN values are within Dempster's bounds, with an uncertainty range of 0.336. In Case 3, the opinion pool and MYCIN values are within Dempster's bounds, with an uncertainty of 0.274. In Case 4, Dempster's rule and the MYCIN method show some reinforcement, although the belief measure $d1$ is a little over the upper bound. The opinion pool values are well outside of Dempster's bounds, with a small uncertainty range of 0.071.

Table 8. Values of $d1$ and $d2$ for decision between $H1$:merchant and $H2$:not merchant.

	CF^{*} measures	Dempster's rule [with prob. bonus]		Linear pool	Log pool	Fuzzy
Case 1						
$H1$	<u>0.745</u>	<u>0.873</u>	[0.745–1.0]	<u>0.646</u>	<u>0.650</u>	<u>0.909</u>
$H2$	0	0.127	[0–0.255]	0.354	0.350	0.786
Case 2						
$H1$	<u>0.625</u>	<u>0.729</u>	[0.561–0.897]	<u>0.573</u>	<u>0.580</u>	<u>0.853</u>
$H2$	0.235	0.271	[0.103–0.439]	0.427	0.420	0.786
Case 3						
$H1$	0.465	0.375	[0.238–0.512]	0.467	0.460	0.787
$H2$	<u>0.640</u>	<u>0.625</u>	[0.488–0.762]	<u>0.533</u>	<u>0.540</u>	<u>0.803</u>
Case 4						
$H1$	0.100	0.044	[0.008–0.079]	0.294	0.276	0.699
$H2$	<u>0.928</u>	<u>0.956</u>	[0.921–0.992]	<u>0.706</u>	<u>0.724</u>	<u>0.877</u>

*Certainty Factor $CF = d1 - d2$

2.2.4 Submarine Innocent Passage

A submarine that is proceeding at normal speed in its usual area of operations or transit is probably in "innocent passage," particularly if it is surfaced. One that is in torpedo range of a high-value target or appears to be hiding by being quiet is of more concern. Table 9 gives probability masses for two submarine cases, each having four pieces of evidence concerning area and behavior. As in the previous example, two components of uncertainty are used in the conversion to $P(Hi|Ej)$. Table 10 gives the results for the four candidate decision methods and for the MYCIN certainty factor method. In Case 1, all favor $H2$ except the fuzzy method, which is close to a tie. The opinion pool and MYCIN values are within Dempster's bounds, with an uncertainty range of 0.290. In Case 2, all methods favor $H1$. The opinion pool and MYCIN values are within Dempster's bounds, with an uncertainty range of 0.24.

Table 9. Evidence and probability masses for decision between $H1$: innocent passage and $H2$: hostile intent, for two submarine cases.

Evidence Ej	$E1$	$E2$	$E3$	$E4$
Case 1	Usual area	Submerged	Quiet	Torpedo range
$mj(H1)$	0.4	0	0	0
$mj(H2)$	0	0.2	0.25	0.4
$mj(U)=mj_E(U)+mj_H(U)$	0.6=0.2+0.4	0.8=0.1+0.7	0.75=0.3+0.45	0.6=0.3+0.3
Case 2	Unusual area	Surfaced	Normal speed	Course change
$mj(H1)$	0	0.6	0.2	0
$mj(H2)$	0.3	0	0	0.3
$mj(U)=mj_E(U)+mj_H(U)$	0.7=0.2+0.5	0.4=0.05+0.35	0.8=0.3+0.5	0.7=0.3+0.4

Table 10. Values of $d1$ and $d2$ for decision between $H1$: innocent passage and $H2$: hostile intent, for two submarine cases.

	CF* measures	Dempster's rule [with prob. bounds]		Linear pool	Log pool	Fuzzy
Case 1						
H1	0.400	0.339	[0.194–0.484]	0.427	0.423	0.706
H2	<u>0.640</u>	<u>0.661</u>	[0.516–0.806]	<u>0.573</u>	<u>0.577</u>	<u>0.699</u>
Case 2						
H1	<u>0.680</u>	<u>0.630</u>	[0.510–0.750]	<u>0.532</u>	<u>0.544</u>	<u>0.744</u>
H2	0.510	0.370	[0.250–0.490]	0.468	0.456	0.600

*Certainty Factor $CF = d1 - d2$

2.3 OBSERVATIONS ON DECISION-MAKING PERFORMANCE

The work described in Section 2 is the first step in finding a satisfactory decision algorithm that uses data-quality measures in situations where numerical probability judgments are limited. Four candidate algorithms were compared in sample problems to see if they give reasonable results. No ground truth is available, so the suitability of an algorithm is in large part determined by how well it agrees with the other methods. Suitability is also determined by the extent to which it satisfies the four desired properties that were specified.

The four candidate algorithms gave surprisingly similar results in these nine examples and in other comparisons not shown. The MYCIN certainty-factor method agreed closely with the four candidates in the six examples where it applies. The first-choice decisions for the two opinion pools were always the same, and usually their ranking throughout was the same or similar. In seven of the nine examples, all methods gave the same first choice. In the other two examples, Dempster's rule gave a different first

choice in one and the fuzzy logic method in the other. In both cases, the first and second choices were close. Several of the examples called for clear-cut decisions, however, and the methods would agree significantly less if all of the examples involved highly conflicting evidence.

The values of the decision statistic *di* for Dempster's rule and the opinion pools sum to 1 and can be treated as estimates of probability. Interestingly, the nonzero values of *di* for the opinion pools were inside Dempster's probability bounds for all 18 hypotheses except the first choice (where the linear pool's *d*8 was low) in table 6 and were outside the bounds for all except the first choice in table 3.

Dempster's rule and the certainty factor method have the reinforcement property, and should not be used on dependent evidence. For example, when all of the evidence favors one hypothesis moderately, the results strongly favor that hypothesis. The opinion-pool values of the statistic *di* tend to fall outside of Dempster's bounds when strong reinforcement occurs, as illustrated in Cases 1 and 4 in table 8.

Conversion between Dempster-Shafer probability assignments and sets of conditional probabilities with weight *Wj* (the measure of evidence quality) was ad hoc. The only purpose of the conversions was to compare the algorithms. In practice, the probability judgments should be elicited in the form required for the method chosen.

Care needs to be taken with prior probabilities, the probabilities of *Hi* given no evidence. Examples here used the base rate (usually having to do with location in a lane or corridor) as one piece of evidence. In many applications, a mechanism for calibrating and using human expectations based on preceding events would also be desirable. Section 4.3 will discuss the problem of eliciting and using probabilities.

The logarithmic opinion pool and the fuzzy method eliminate consideration of a valid hypothesis *Hi* when bad evidence yields $P(Hi|Ej) = 0$ or $P(Ej|Hi) = 0$; i.e., they fail Property 3. Table 11 gives a simple example where their use is not advised. The evidence consists of three intercepted radar signals. (In practice, a base-rate distribution is also needed, to indicate the number of ships of each class normally in that area.) The probabilities $P(Hi|Ej)$ are 1/10, 1/7, and 1/2, respectively, for hypotheses having radars x, y, and z. The hypotheses correspond to ship classes, only one of which carries all three radars. (This example is discussed further in Section 4.1.2.) Note that if $Wj = 1$ for one of the intercepted radar signals, the linear pool does not necessarily reject the hypotheses for classes not carrying that radar, thereby it fails Property 1.

None of the four decision algorithms is fully satisfactory. Dempster's rule operates only on independent evidence. Also, the domain expert may find it difficult to properly represent uncertainty and subjective beliefs as probability mass assignments. The linear opinion pool does not satisfy Property 1; that is, it does not necessarily reject hypothesis Hi when $P(Hi|Ej) = 0$ and the evidence Ej is absolutely correct. Neither opinion pool satisfies Property 2; i.e., Hi is not always selected when $P(Hi|Ej) = 1$ and the evidence is absolutely correct. As discussed just above, the logarithmic opinion pool and the fuzzy method fail Property 3. An algorithm is needed that has the desired properties and employs measures of correlation or dependence among the evidence.

Table 11. Comparison of decisions for three-emitter example.

Evidence: $E1$: radar x $W1 = 0.8$
$E2$: radar y $W2 = 0.8$
$E3$: radar z $W3 = 0.4$

Hypothesis (Ship class)	Radars possessed			Dempster's rule		Linear pool	Log pool and Fuzzy
$H1$	x			0.042	[0.038–0.131]	0.040	0
$H2$	x			0.042	[0.038–0.131]	0.040	0
$H3$	x			0.042	[0.038–0.131]	0.040	0
$H4$	x			0.042	[0.038–0.131]	0.040	0
$H5$	x			0.042	[0.038–0.131]	0.040	0
$H6$	x			0.042	[0.038–0.131]	0.040	0
$H7$	x			0.042	[0.038–0.131]	0.040	0
$H8$	x		z	0.086	[0.081–0.175]	0.140	0
$H9$	x	y	z	0.186	[0.181–0.275]	0.197	1
$H10$	x	y		0.117	[0.113–0.206]	0.097	0
$H11$		y		0.058	[0.054–0.147]	0.057	0
$H12$		y		0.058	[0.054–0.147]	0.057	0
$H13$		y		0.058	[0.054–0.147]	0.057	0
$H14$		y		0.058	[0.054–0.147]	0.057	0
$H15$		y		0.058	[0.054–0.147]	0.057	0
$H16$				0.0047	[0–0.094]	0	0
$H17$				0.0047	[0–0.094]	0	0
$H18$				0.0047	[0–0.094]	0	0
$H19$				0.0047	[0–0.094]	0	0
$H20$				0.0047	[0–0.094]	0	0

3.0 MEASURES OF INFORMATION VALUE AND CONFLICT

If the decision algorithm is used in a knowledge-based system, statistical techniques can be combined with reasoning techniques to select relevant data and to identify bad or conflicting data. First, we consider ways of measuring the relevance of data.

3.1 MEASURES OF RELEVANCE

Evidence that is likely to occur only with one hypothesis is highly relevant. Figure 3 gives an example for four hypotheses. (Equation 2 gives the relationship between $P(Ej|Hi)$ and $P(Hi|Ej)$.) Evidence that rules out an important hypothesis is also of value. A piece of evidence is irrelevant to the decision problem if it is likely to occur for all hypotheses, that is, if $P(Ej|Hi)$ is large for all Hi. Figure 4 illustrates such a case. Note that when represented in the form $P(Hi|Ej)$, we lose the information about Ej being highly likely. However, evidence is not useful if it is nearly equally likely for all hypotheses, which includes the latter case.

One function we can use to measure irrelevance in the above sense is the entropy function

$$\text{Irrelevance} = -\sum_i P(Hi|Ej) \log P(Hi|Ej), \tag{15}$$

which is maximized (with value $\log n$) when $P(Hi|Ej) = 1/n$ for all i, i.e., when the hypotheses are equally likely. A slightly simpler measure is the function

$$\text{Relevance} = -\sum_i \log P(Hi|Ej), \tag{16}$$

which is minimized (with value $n \log n$) in the equal likelihood case. Neither of these functions is defined if $P(Hi|Ej) = 0$ for some i. A simple function that will work in all cases is

$$\text{Relevance} = \max_i P(Hi|Ej) - \min_i P(Hi|Ej). \tag{17}$$

Computations for many probability distributions showed all three functions to give the same ordering, although some ties occurred with the third.

One approach to measuring how well a parameter measurement contributes toward discriminating between two hypotheses is to use distance measures. Figures 5 and 6 give examples of applications. While a middle value of the parameter is much less useful than an extreme value, the distance measure gives a measure of the overall benefit of a parameter to that particular decision problem. Whether the measurement is actually used for that problem would depend on its value.

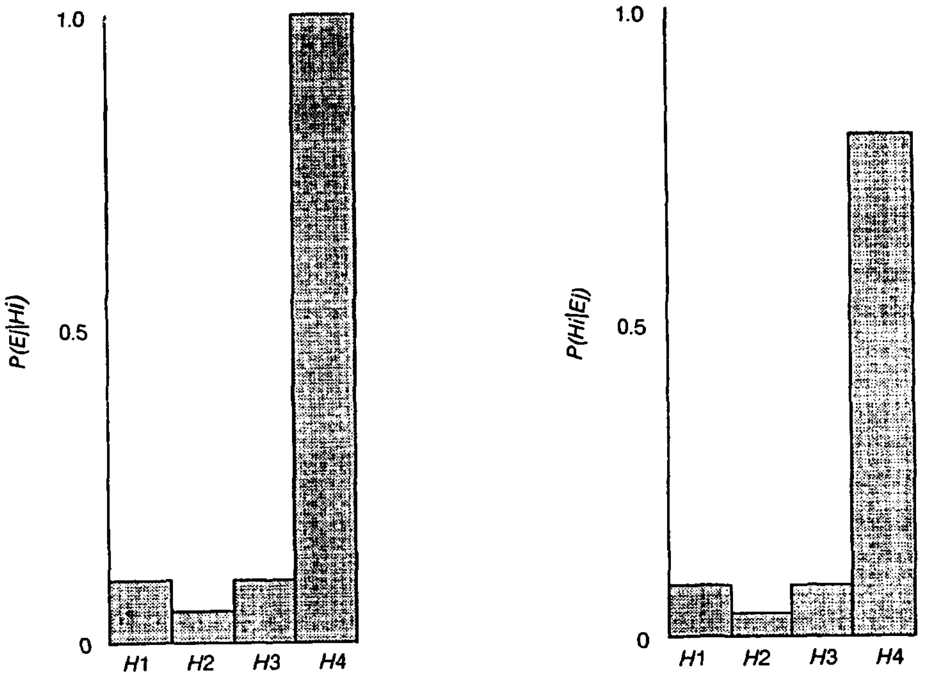

Figure 3. Evidence highly unique to a hypothesis.

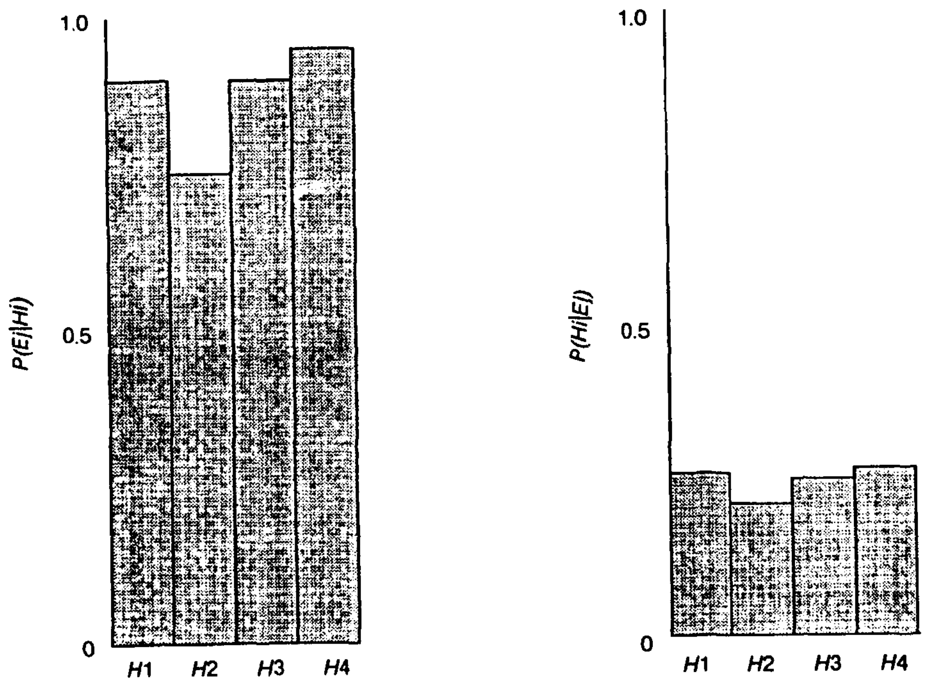

Figure 4. Evidence likely for any hypothesis.

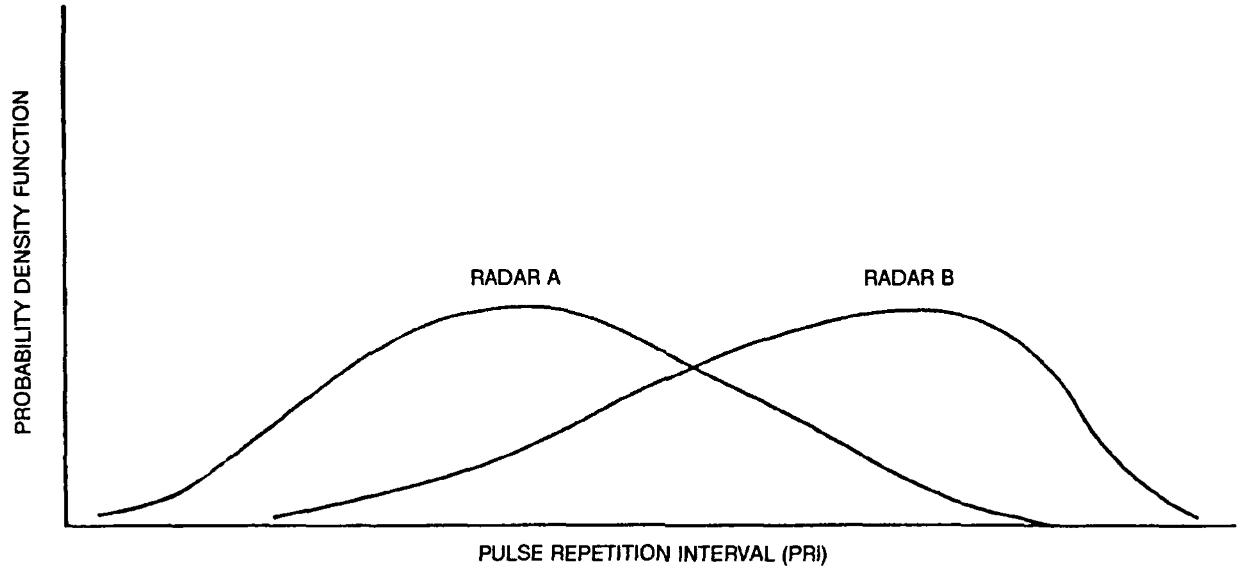

Figure 5. Distance measure application—information value of PRI measurement for distinguishing between radars having PRI agility.

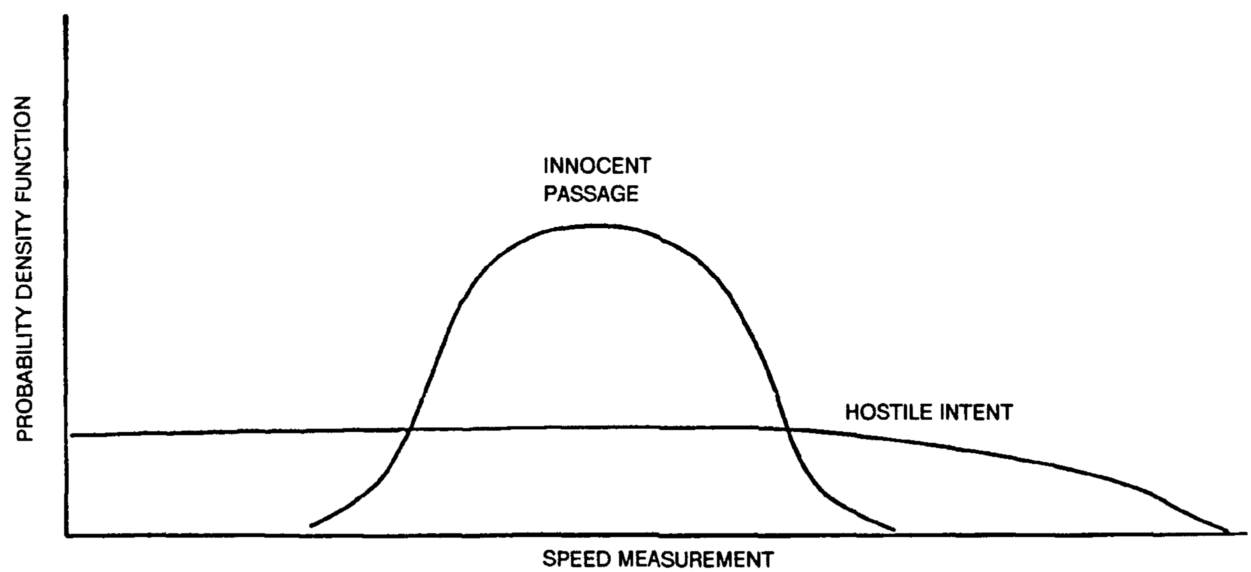

Figure 6. Distance measure application—information value of single-speed measurement for deciding submarine intent.

The following are measures of the distance between the conditional densities, $p_1(x)$ and $p_2(x)$, of two hypotheses (Lainiotis & Park, 1972). As we will see in another application in the next section, they give similar results.

Kullback Divergence:

$$J = \int [p_1(x) - p_2(x)] \log[p_1(x)/p_2(x)]dx \tag{18}$$

Bhattacharyya Distance:

$$\mathrm{B} = -\log \rho \tag{19}$$

where ρ is the Bhattacharyya coefficient

$$\rho = \int \sqrt{p_1(x)p_2(x)}\, dx \tag{20}$$

Kolmogorov Variational Distance:

$$K = \int |p_1(x) - p_2(x)|\, dx \tag{21}$$

Matusita Distance:

$$M = \{\int [\sqrt{p_1(x)} - \sqrt{p_2(x)}]^2\, dx\}^{1/2} \tag{22}$$

Lainiotis and Park (1972) point out that the Bhattacharyya coefficient, ρ, is related to the Matusita distance, M, by

$$M = \sqrt{2(1 - \rho)} \tag{23}$$

and to the Kullback divergence, J, by the inequality

$$\rho \geq \exp(-J/4\}. \tag{24}$$

3.2 MEASURES OF CONFLICT

Questionable data should be called to the attention of the operator or commander, who can have it verified or downgrade its credibility. Evidence Ej is suspicious if the probability $P(Ej|Hi)$ is low for all Hi (as it was for $E4$ in the Iranian airbus example in Section 2.2.1). Figure 7 illustrates such a case. Note that this information is not available if the distribution is in the form $P(Hi|Ej)$. Also, evidence is sometimes questionable if it gives a probability distribution very different from those for other evidence. A rule-based system could determine if the disagreement is normal for that type of problem and kind of evidence. For example, the distributions for the three-emitter case in Section 2.3 would show a large measure of disagreement. Next, we investigate methods of determining when the distribution contributed by one piece of evidence differs significantly from the others.

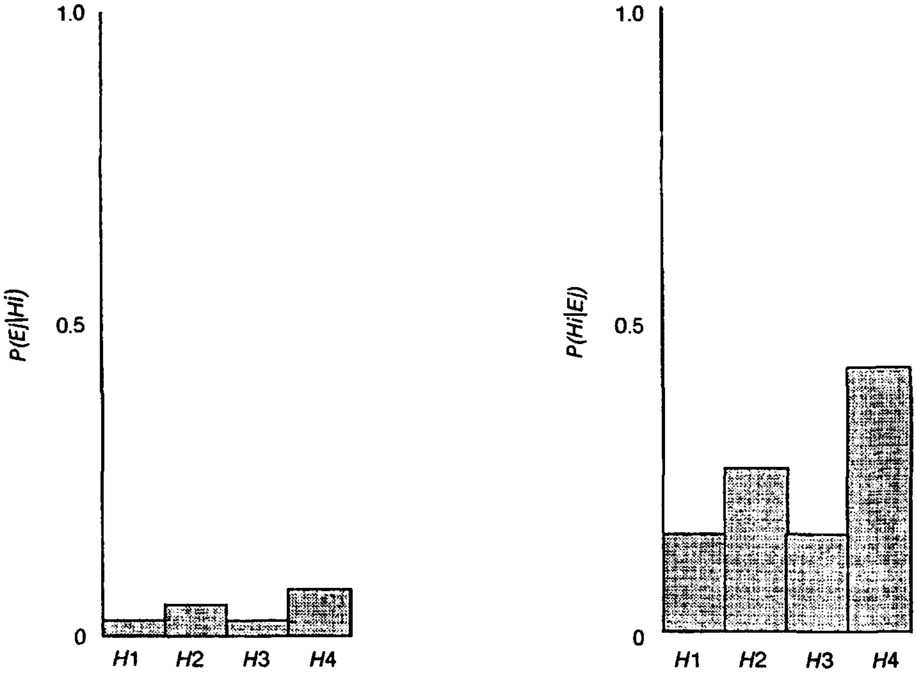

Figure 7. Evidence unlikely for any hypothesis.

3.2.1 Dempster Assignments

The measure of conflict between two assignments *mj* and *mk* is

$$\text{conflict}(Ej,Ek) = \sum_{Hi \& H1 = \emptyset} mj(Hi) \bullet mk(H1). \tag{25}$$

This quantity is $1 - C$, where C is the normalizing constant given by Equation 7. Figure 8 illustrates this area of conflict for two very different distributions. Note that in figure 9, however, two identical distributions also show a considerable amount of conflict. Figure 10 illustrates a case where no conflict occurs. (Equation 25 does not apply to assignments giving probability to disjunctions of hypotheses other than *U*.) We can also compute the conflict among all distributions, but here we wish to determine if one distribution conflicts with the others significantly more. We use Equation 25 for $k < j$ and let the conflict for an individual *Ej* be

$$\text{conflict}(Ej) = \sum_k \text{conflict}(Ej,Ek) + \sum_k \text{conflict}(Ek,Ej). \tag{26}$$

When computing conflict for the purpose of determining if one distribution is very different from the others, we should not include the uncertainty about the data in the

uncertainty measures. For comparisons with the distance measures below, we let $mj(U)$ be 0 for all j; i.e., we let $mj(Hi) = P(Hi|Ej)$. For this case, Equation 25 becomes

$$\text{conflict}(Ej,Ek) = 1 - \sum_i P(Hi|Ej) \bullet P(Hi|Ek) \ . \tag{27}$$

When the distributions are identical, we have

$$\text{conflict}(Ej,Ek) = 1 - \sum_i [P(Hi|Ej)]^2 \ , \tag{28}$$

which is maximized with value $1 - 1/n$ when $P(Hi|Ej) = 1/n$ for all i, i.e., when the hypotheses are equally likely. We further note from Equation 27 that when Ej gives an equally likely distribution and Ek gives a different distribution, we also have conflict $(Ej,Ek) = 1 - 1/n$. (While the conflict between the two distributions is greater, the internal conflict of the distribution from Ek is less.) We then see that this measure of conflict is unsuitable for our purpose, but it is of interest to compare it with other measures.

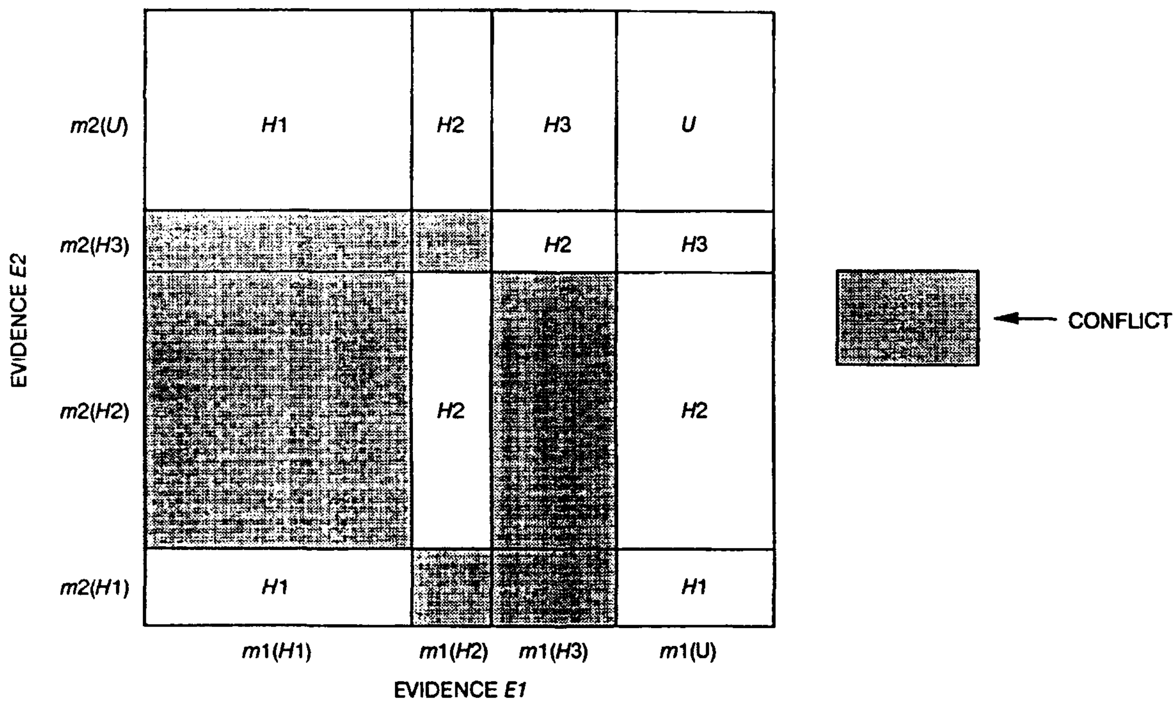

Figure 8. Combination of conflicting evidence by Dempster's rule.

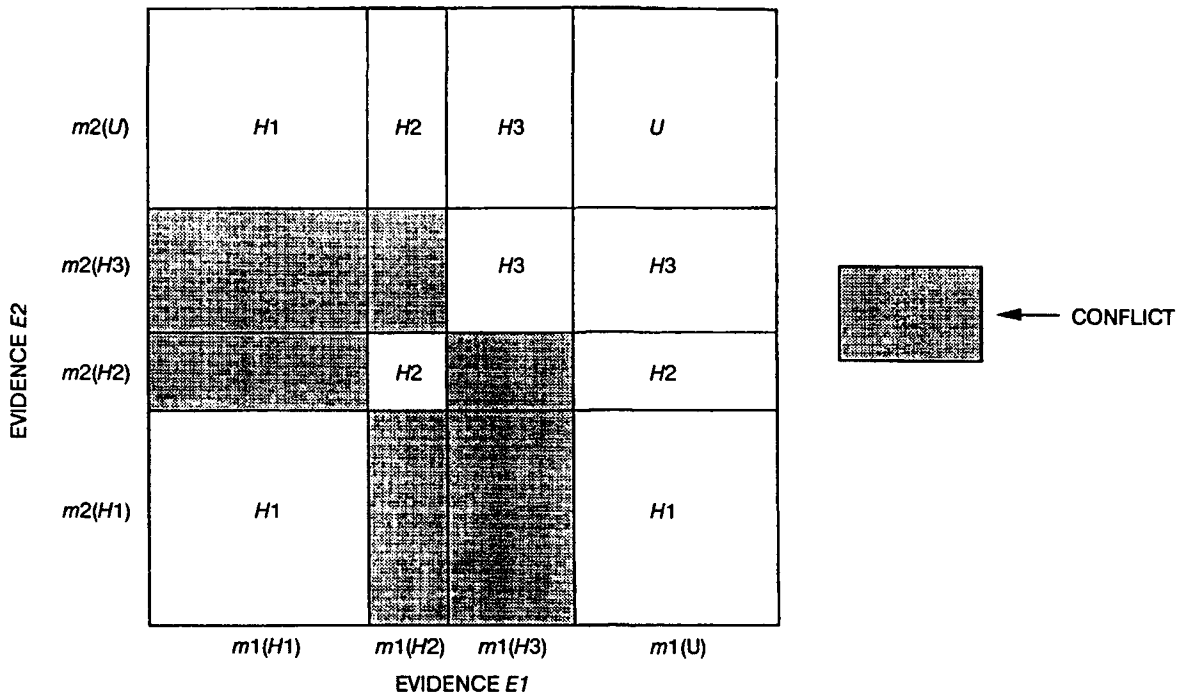

Figure 9. Combination of consistent evidence by Dempster's rule.

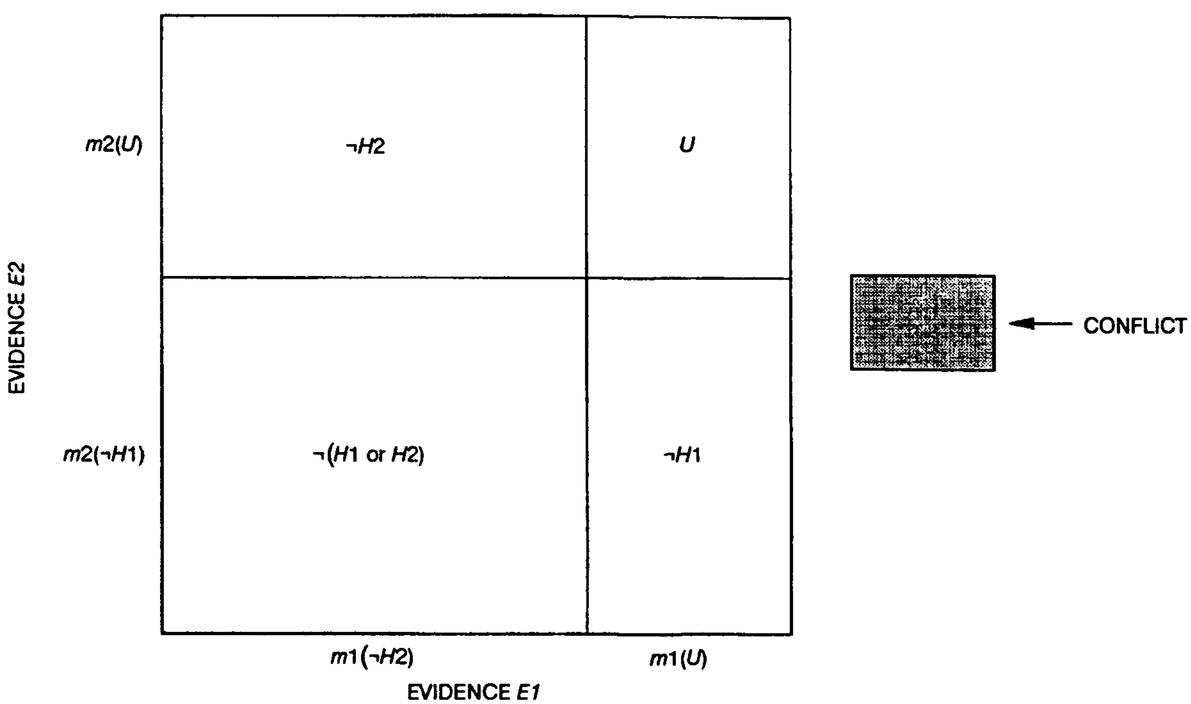

Figure 10. Combination of nonconflicting evidence by Dempster's rule (number of hypotheses >2).

3.2.2 Probability Distance Measures

For distributions of the form $P(Hi|Ej)$, we can use the following discrete versions of the pairwise distance measures given in the last section.

$$\underline{\text{Kullback}}\text{: distance}(Ej,Ek) = \sum_i \left[P(Hi|Ej) - P(Hi|Ek)\right] \log \frac{P(Hi|Ej)}{P(Hi|Ek)} \tag{29}$$

$$\underline{\text{Bhattacharyya}}\text{: distance}(Ej,Ek) = \log\left[\sum_i \sqrt{P(Hi|Ej) \bullet P(Hi|Ek)}\right] \tag{30}$$

$$\underline{\text{Kolmogorov}}\text{: distance}(Ej,Ek) = \sum_i \left|P(Hi|Ej) - P(Hi|Ek)\right| \tag{31}$$

$$\underline{\text{Matusita}}\text{: distance}(Ej,Ek) = \left\{\sum_i \left[\sqrt{P(Hi|Ej)} - \sqrt{P(Hi|Ek)}\right]^2\right\}^{1/2} \tag{32}$$

The distances are computed for $k < j$ and are combined by using the formula

$$\text{distance}(Ej) = \sum_k \text{distance}(Ej,Ek)) + \sum_k \text{distance}(Ek,Ej)) \tag{33}$$

3.2.3 Comparisons

Here we compare the measure (Equation 26) based on Dempster's measure of conflict (Equation 26 or 27) and the measure (Equation 33) based on the four distance measures (Equations 29 through 32) in four examples.

Data Set 1 is shown in figure 11 and listed in table 12. The results are given in table 13, where the ordering is "1" for the greatest conflict and "4" for the least. The four distance measures give very similar results. The Dempster method agrees with the other methods that $E2$ has the greatest conflict, but its second choice, $E3$, has the least conflict according to the other methods.

Data Set 2 is shown in figure 12 and listed in table 14. The results in table 15 show all methods to give essentially the same results.

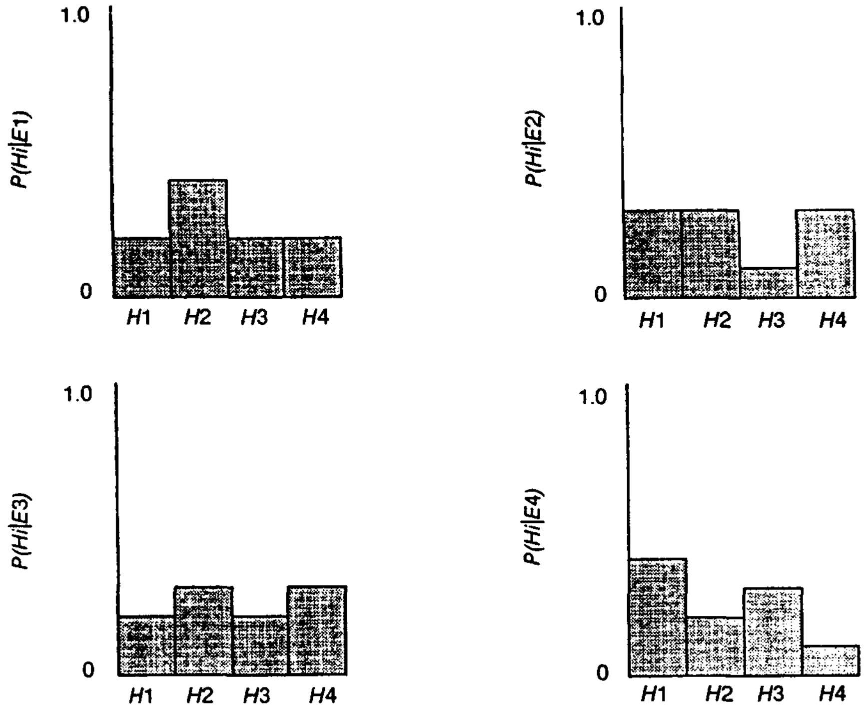

Figure 11. Data Set 1.

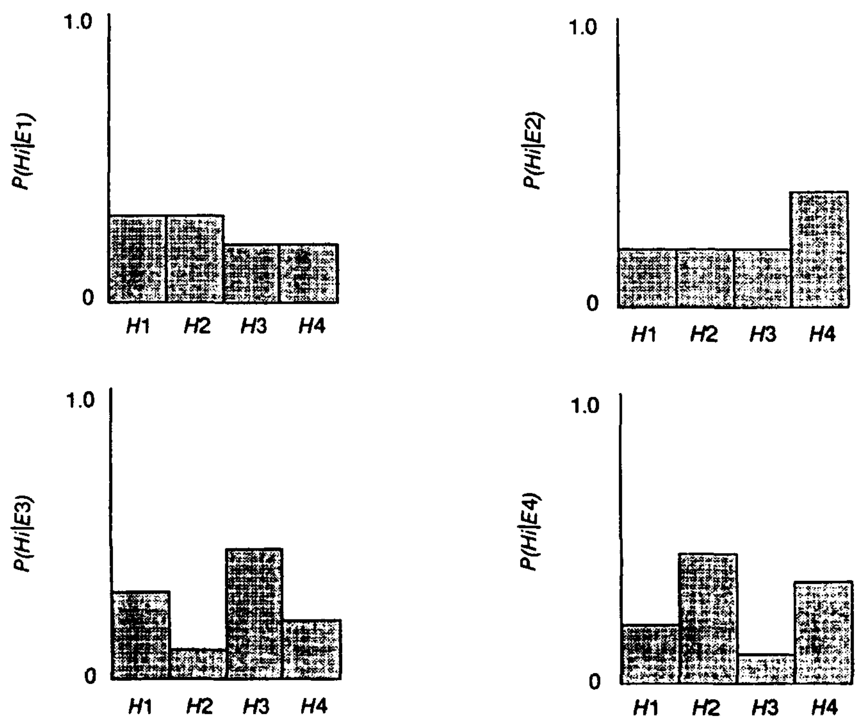

Figure 12. Data Set 2.

Table 12. Data Set 1.

	$P(Hi\|Ej)$			
	$E1$	$E2$	$E3$	$E4$
$H1$	0.2	0.3	0.2	0.4
$H2$	0.4	0.3	0.3	0.2
$H3$	0.2	0.1	0.2	0.3
$H4$	0.2	0.3	0.3	0.1

Table 13. Measures of distance for Data Set 1.

	$E1$	$E2$	$E3$	$E4$
Kullback	0.636 3	0.798 2	0.619 4	1.335 1
Bhattacharyya	0.803 3	0.100 2	0.078 4	0.169 1
Kolmogorov	1.200 2/3	1.200 2/3	1.000 4	1.800 1
Matusita	0.652 3	0.729 2	0.625 4	0.991 1
Dempster	2.240 3/4	2.240 3/4	2.250 2	2.290 1

Table 14. Data Set 2.

	$P(Hi\|Ej)$			
	$E1$	$E2$	$E3$	$E4$
$H1$	0.3	0.2	0.3	0.2
$H2$	0.3	0.2	0.1	0.4
$H3$	0.2	0.2	0.4	0.1
$H4$	0.2	0.4	0.2	0.3

Table 15. Measures of distance for Data Set 2.

	$E1$	$E2$	$E3$	$E4$
Kullback	0.757	0.844	1.658	1.329
	4	3	1	2
Bhattacharyya	0.095	0.106	0.211	0.169
	4	3	1	2
Kolmogorov	1.200	1.400	1.800	1.600
	4	3	1	2
Matusita	0.741	0.785	1.075	0.922
	4	3	1	2
Dempster	2.260	2.260	2.320	2.280
	3/4	3/4	1	2

Data Set 3 is shown in figure 13 and listed in table 16. Note in table 17 that the distribution contributed by $E1$ is much more extreme than the others. $E1$ is given the greatest conflict by the four distance measures but the least by the Dempster approach. This is because the internal conflict of $\{P(Hi|E1)\}$ is much smaller than for the others (see Section 3.2.1).

Table 16. Data Set 3.

	$P(Hi\|Ej)$			
	$E1$	$E2$	$E3$	$E4$
$H1$	0.01	0.1	0.05	0.2
$H2$	0.01	0.2	0.25	0.05
$H3$	0.01	0.3	0.4	0.35
$H4$	0.97	0.4	0.3	0.4

Data Set 4, shown in figure 14 and given in table 18, contains zeroes, so Kullback's distance measure cannot be used. In table 19, the other three distance measures give the same ordering and the Dempster method gives somewhat similar results.

As pointed out in Section 3.2.1, the Dempster measure of conflict is unsuitable for this purpose, although it gives results somewhat similar to the distance measures in these examples. The distance measures all give very similar results. We conclude that the best approach is to use the Kolmogorov distance, because it is the simplest to compute.

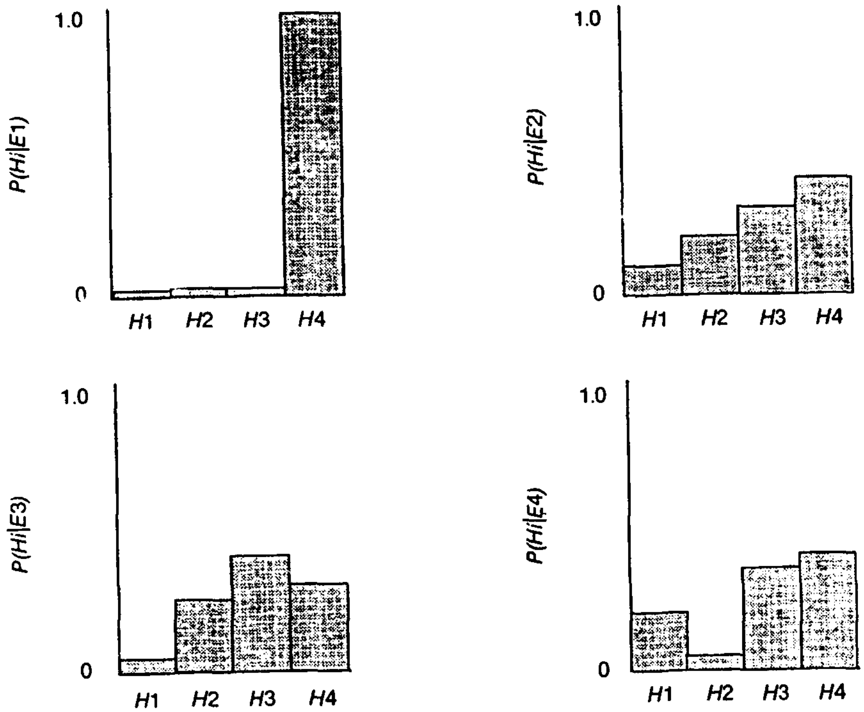

Figure 13. Data Set 3.

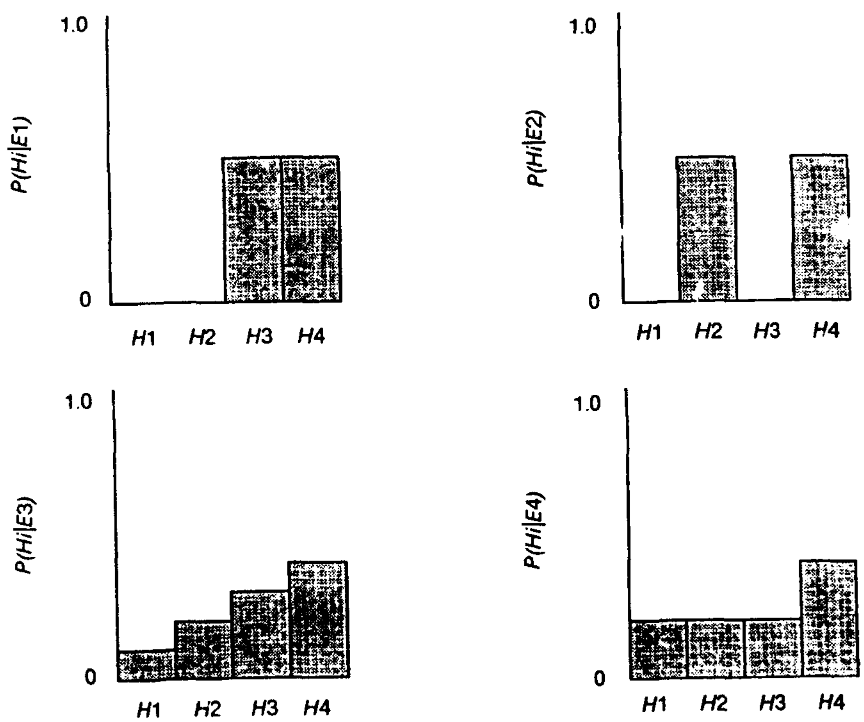

Figure 14. Data Set 4.

Table 17. Measures of distance for Data Set 3.

	$E1$	$E2$	$E3$	$E4$
Kullback	7.677 1	2.656 4	3.705 2	3.198 3
Bhattacharyya	0.964 1	0.330 4	0.476 2	0.394 3
Kolmogorov	3.620 1	1.740 4	2.140 2	1.940 3
Matusita	2.216 1	1.125 4	1.334 3	1.339 2
Dempster	1.914 4	2.016 3	2.125 1	2.028 2

Table 18. Data Set 4.

			$P(Hi\|Ej)$	
	$E1$	$E2$	$E3$	$E4$
$H1$	0	0	0.1	0.2
$H2$	0	0.5	0.2	0.2
$H3$	0.5	0	0.3	0.2
$H4$	0.5	0.5	0.4	0.4

Table 19. Measures of distance for Data Set 4.

	$E1$	$E2$	$E3$	$E4$
Kullback	Not defined for probabilities that are zero			
Bhattacharyya	1.144 2	1.233 1	0.465 4	0.554
Kolmogorov	2.400 2	2.600 1	1.600 4	1.800 3
Matusita	2.263 2	2.376 1	1.428 4	1.541 3
Dempster	2.100 3	2.150 1	2.070 4	2.120 2

4.0 KNOWLEDGE-BASED IMPLEMENTATION

An approach to implementing the decision process in an operational system is outlined here. The decision algorithms and a simple version of a rule-based approach were programmed in C-Language Integrated Production System (CLIPS), a widely used system developed at the NASA/Johnson Space Center. The problem of obtaining the necessary numerical probability judgments is also addressed.

4.1 DECISION PROBLEM ORGANIZATION

4.1.1 Unknown Air Contact Example

Figure 15 illustrates a way of organizing evidence when an unidentified aircraft has originated from a nearby airfield in a country not considered a friend The output of each box is a numerical probability assignment and is shown in the form of CLIPS facts. The assignment is either in the form (*p* evidence hypothesis number time), which represents $P(Hi|Ej)$, or in the form (*pe* evidence hypothesis number time), which represents $P(Ej|Hi)$. The latter assignments will often be derived as fuzzy measures of the consistency of the evidence with the hypothesis.

For many kinds of evidence, we can choose either to use rules to create the numerical assignments as needed or to store and maintain assignments in a system database for all possible values of the evidence. The choice will depend on the efficiency of each method in the knowledge-based shell used. For experimental purposes, the assignments can be read from files.

The first box in figure 15 generates probabilities very much like priors, using the base rate (for the three aircraft uses, based on origin) and a measure of the expectation of hostilities. (We avoided the issue of priors in Section 2.2.1, and implicitly assumed that all hypotheses were equally likely.) The rate for military aircraft can be divided between $H1$ and $H2$ based on this expectation, which is an estimate of the conditional probability $P(\text{hostile}|\text{military})$. Ideally, this estimate is predetermined. However, except in a steady-state political environment, a commander's interpretation of very recent events will be needed to continually update its value. (The subject of using human expectations is discussed in more detail in Section 4.3.2.) In some situations, every military aircraft from an unfriendly air field will be considered hostile. Distinguishing between the uncertainty concerning evidence and uncertainty in the interpretation of the evidence is difficult when dealing with $H1$ and $H2$. Uncertainty about the estimated probability of hostilities usually would be larger than the uncertainty about base rate, and the quality measure $W1$ should reflect both uncertainties.

The latitude and longitude information may indicate a classic commercial aircraft takeoff or may be consistent with the common behavior of other users of the airfield. Computations would also measure adherence to the centerline of the commercial

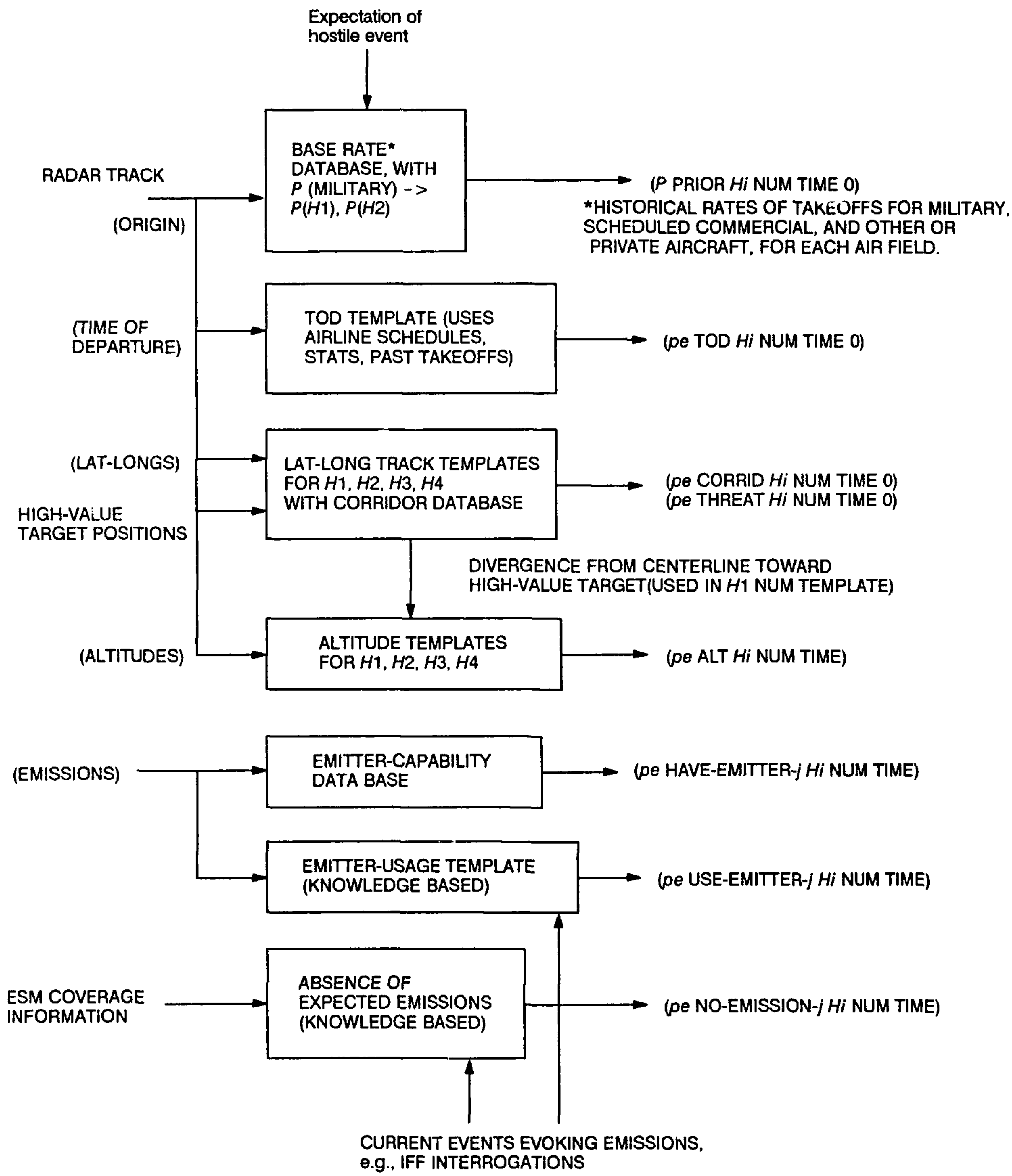

Figure 15. Generation of conditional probabilities for unknown air contact, airfield takeoff case. ($H1$:military-hostile intent, $H2$:military-innocent transit, $H3$:commercial carrier, $H4$:private or other commercial).

corridor after takeoff. $H3$ will receive a large probability if the track is typical of commercial takeoffs and $H1$ will receive a large probability if the aircraft is off-center and flying directly toward high-value target. Decreasing altitude indicates hostility if the latter is true. Altitude information is treated separately from other position information partly because the accuracy measures will usually be different. If a sophisticated template is designed to compare three-dimensional tracks with historical tracks, the accuracies could be combined and a single box would replace the lat-long box and altitude box.

In our simple rule-based experimental version, the probability assignment for adherence to a corridor is a function of the contact's distance from the center-line and its angle divergence from the centerline. Information about earlier behavior is used when this assignment is updated. A threat value based on distance and direction is computed for each high-value target, and the assignment is based only on the target with the greatest threat. A default assignment occurs if there is no threat. The altitude is considered separately. $H1$ receives the larger probability if the altitude is not high and is decreasing.

The time of departure (TOD) template will most affect the estimate of $P(TOD|H3)$, since only commercial carriers have schedules. Night takeoffs may be unlikely for $H4$.

The bottom box in figure 15 typically would correspond to IFF mode or to non-response to a radio warning. Other evidence not shown could include specific reactions or nonreaction to a fire-control radar.

Although not shown, the CLIPS facts also need an element giving the track number or other identification code. Simultaneous processing of all current contacts on a parallel computer would avoid having to include that element and would be much faster.

Special rules are needed to note when the probability of certain evidence occurring is small for all hypotheses, which indicates the evidence may be in error. Other rules could determine when the different pieces of evidence give very different probabilities.

4.1.2 Ship-Type Example

In addition to the prestored distributions of the kind used in the example in Section 2.2.2, templates can be used for interpreting changes in course and speed, etc., as in figure 15. Another kind of evidence is the contact's position and course relative to an earlier sighted ship whose type was known. That evidence is easy to use with Dempster's rule, but awkward with the other methods.

Emissions from the contact can be used in a separate decision process, described in Section 2.3 and shown in table 11. To combine emission results with the other

evidence on ship type, the base rate must be combined with the emission distributions at this class level and not at the type level. For efficiency, match class hypotheses to those classes having at least one of the emitters and to super classes that cover all other classes. e.g., $H16$ may be the super class of all other classes or types not having x, y, or z, of the nationality corresponding to the emitters, and $H17$ may be all U. S. military ships, and so forth. The base rate is then used as another piece of evidence, and the rate for the super classes will generally far exceed those for specific classes. The resulting values of di for individual types in $H16$, etc., can then be found by dividing up di in proportion to their base rate. The resulting class values of di are then summed over each hypothesized type, and that distribution is used with the others in the contact-type decision process. Rather than throw away valuable information, the types containing classes having radars x, y, or z should be divided into two sub-hypotheses, one for that nationality and one for all others. Implementation of this scheme would be complex, but would certainly be feasible in a rule-based system.

4.2 ASSIGNMENT TO DISJUNCTIONS FOR DEMPSTER'S RULE

There are often advantages to using the general form of Dempster's rule. In a decision about ship type, behavior unlike a merchant would assign probability to $\neg H15$, i.e., to Not Merchant. If the contact's position is within reach of an earlier sighted platform whose type was known, probability need only be assigned to that ship type and to U (the disjunction of all hypotheses). In the unknown-air case, origin would assign probabilities to $H3$ and $H4$ and to the disjunction of $H1$ and $H2$ ($H1$v$H2$) rather $H1$ and $H2$ individually. Capability of IFF would also assign probability to the disjunction of $H1$ and $H2$. Then, the a priori probability of hostile action (conditional on the contact being military) is not needed to divide the probability between $H1$ and $H2$. Evidence suggesting innocent transit would assign mass to $\neg H1$, the disjunction of $H2$, $H3$, and $H4$. A departure when none had been scheduled for a while would give probability to $\neg H3$. The distributions in all these cases are of the form (p Ej B number time), where B may be Hi, $\neg Hi$, U, or a disjunction such as $H2$v$H6$v$H12$.

The combination of assignments for this general case is much more computationally intensive than the method in Section 2.1.4, and the decision statistic di would have to be defined such that the values sum to 1. Equation 9 applies, except that the uncertainty term $m(U)$ is also a function of i.

The greatest difficulty in eliciting assignments occurs when subjective judgments are made on a large number of hypotheses. For example, the expert might say, "I'm 60% sure it's not A3 and 70% sure it's not A17." These are "support" values, where the support for a general proposition B is (Shafer, 1976)

$$S(B) = \sum_{B' \& B = B'} m(B')$$

The mass assignment for this example is $m(\neg(A3 \vee A17)) = 0.6$, $m(\neg A17) = 0.1$, and $m(U) = 0.3$. With some training, a user can quickly convert his specified support values into probability mass values. In general, a specialized intelligent interface would be needed for eliciting assignments from multiple domain experts.

4.3 ELICITING AND USING EXPERTS' JUDGMENTS

Some of the distributions $\{P(Hi|Ej)\}$ or other numerical probability judgments will come from empirical data and some from subjective judgments by experts. We discussed empirical methods in Section 2.2; here we consider how to elicit and aggregate probabilities from experts. Much of the discussion on expert estimates and aggregation in Section 4.3.1 applies also to eliciting expectations of hostilities, discussed in Section 4.3.2.

4.3.1 Eliciting Probabilities

Experts often disagree on their estimates of a probability, sometimes because they make much different assumptions. They can either be asked to agree among themselves on a single probability or probability distribution or they can give different answers that need to be mathematically aggregated. If the problem does not involve deep problem solving, a combination of their estimates is generally better than the *estimate from best expert* (Meyer & Booker, 1991). The median of the expert's estimates is a commonly used aggregation estimator. (This is satisfactory for estimates of $P(Ej|Hi)$ or of the probability of hostilities, but estimates of $P(Hi|Ej)$ would have to be normalized to sum to one.) Other popular estimates are the mean and the geometric mean. Using the mean has the advantage of an easily calculated variance. The three estimates (median, mean, and geometric mean) can give very different results when the probabilities are for a very unlikely event. A median or geometric mean estimate works better for this case (Meyer & Booker, 1991).

Many experts prefer to give a range of possible values of a probability instead of a single point estimate. (Ranges would be appropriate for $P(Ej|Hi)$.) Their range of uncertainty could be, for example, the 40th or 60th percentile values, fractions of sigmas, or multiples of sigmas. However, Meyer and Booker (1991) point out that experts underestimate uncertainty. The percentiles given really represent a fraction of the true uncertainty: People estimate the 60 to 70th percentile values when asked for the 95th, and the 30 to 40th percentile values when asked for the 5th. If the experts provide ranges and these ranges do not overlap, the possibility that their interpretation of the question or their assumptions differ should be explored. When they do overlap, the aggregation method should select a point within the overlapped area.

If the individual or group uncertainty is significant, a decision method (such as Dempster's rule) that uses a measure of the uncertainty would be preferred over the

other methods. If ranges are given, an estimate of the uncertainty $m(U)$ can be derived from the range among experts or from individual expert's ranges. The uncertainty measure would need to be across all hypotheses. When point estimates are given, the variances can be used to estimate the uncertainty. Another possibility is to use distance measures. Distance measures can be used to compare distributions among experts for one kind of evidence in exactly the same way distributions among evidence were compared in Section 3.2.

An alternative to asking the expert for the numerical values of a probability distribution is to provide a graphical interface. The user could increase and decrease elements of histograms such as those in figure 4. (If a probability is close to 0 or 1, the number itself would need to be typed in.) When estimating $\{P(Hi|Ej)\}$, the program would maintain a unity sum over i.

To the extent possible, online documentation of the method or reasoning used to create a probability distribution should be easily accessible to the user. Some of the distributions may be created from numerical evidence by using formulas agreed on by the experts. Some of the look-up distributions may have to be frequently updated. The knowledge-based system should send time-to-update alerts and provide tools for reviewing the situation and making changes.

4.3.2 Expectations of Hostile Action

In the upper box of figure 15, the "expectation of a hostile event" is used to divide the base-rate probability for military aircraft into two probabilities, one for the hypothesis of hostile military and one for the hypothesis of military in innocent transit. The first of these probabilities is given by

$$P(\text{hostile military}) = P(\text{hostile}|\text{military})\ P(\text{military})$$

and the other is the difference between this and the probability that the contact is military. The conditional probability used in this equation must be carefully explained when elicited: If the aircraft departing that particular airfield (or in that particular corridor) is military, what is the probability the intent is hostile, independent of any other observations of the aircraft. If a hostile action is subsequently taken, there generally will be a state change, and this probability will become large. The initial probability will be based largely on recent political or military activities. The expert's estimate or the experts' aggregated estimate of this conditional probability is the best knowledge available. However, we should note that humans tend to overestimate the likelihood of rare events (Meyer & Booker, 1991). Also, experts directly threatened can be expected to give higher probabilities.

Unless a good estimate of this conditional probability is available, it is better to avoid using it. One way is to assign $P(\text{military})$ to the disjunction of the two

hypotheses, and use the general case of Dempster's rule. Often, such as in the submarine example in Section 2.2.4, there is no base rate to divide, and one has the option of using an a priori probability of hostile intent. In that example, there are only two hypotheses, so the a priori distribution consists of that probability and its complement. In some cases, the hypotheses will correspond to various kinds of hostile actions that could be taken.

The main product of the knowledge-based system would be a numerical or graphical presentation based on the results of the decision algorithm. If an a priori probability of hostile action is used, it should be made clear to the operator or commander receiving the decision results. When one of the available inputs is that probability (or set of probabilities), the three alternatives are to (1) present the integrated results, (2) present two sets of results, the integrated one and one without (i.e., based only on direct observations of the contact), and (3) present the latter. In all three cases, the a priori probability or distribution can be presented simultaneously. (Some experts may prefer to choose among judgments ranging from Highly Possible to Can Never Happen, in which case option 3 is the obvious choice.) An interesting study would be to find the sensitivity of the decision results to a priori probabilities.

5.0 CONCLUSIONS

The first problem addressed was to find a decision method for deciding among a set of hypotheses $\{Hi\}$ based on a set of evidence $\{Ej\}$, given data-quality factors $\{Wj\}$ and numerical-probability judgments. The latter are typically of the form $P(Ej|Hi)$ or $P(Hi|Ej)$. Of the four algorithms compared in Section 2, none was satisfactory for all decision problems. Dempster's rule operates only on independent evidence. The linear opinion pool does not necessarily reject hypothesis *Hi* when $P(Hi|Ej) = 0$ and the evidence *Ej* is absolutely correct. For both the linear and logarithmic opinion pools, *Hi* is not always selected when $P(Hi|Ej) = 1$ and the evidence is absolutely correct. The logarithmic opinion pool and the fuzzy method eliminate consideration of a valid hypothesis Hi when bad evidence yields $P(Hi|Ej) = 0$ or $P(Hi|Ej) = 0$. For most of the examples considered, the four methods gave similar results. If selected carefully on the basis of the decision problem, at least one of these methods should perform adequately. As pointed out in Section 2.3, however, a method is needed that has the desired properties and employs measures of correlation or dependence among the evidence.

Several statistical techniques were investigated for measuring the relevance of evidence to a decision problem and for identifying *suspicious or conflicting* evidence. Relevance measures were described in Section 3.1 and conflict measures in Section 3.2. The measures of distance between two distributions were shown to be applicable to both problems. The measures are also applicable when aggregating probability distributions from a group of experts, as pointed out in Section 4.3.

Is it reasonable to expect that we can successfully implement these statistical techniques? Section 4.1 describes an approach for two representative problems. A simple version of one was implemented in CLIPS to uncover unforeseen difficulties that could occur. (An operational system should employ specialized event templates to best exploit all information about the contact's behavior in the context of the larger picture.) The approach seems feasible, although it depends on having current probability assignments or other numerical judgments from experts. Sometimes an estimate of the a priori probability of an event is needed, e.g., $P(\text{hostile}|\text{military})$. Section 4.2 suggests ways to avoid having to use this probability by assigning probability to disjunctions of hypotheses and using the general form of Dempster's rule; however, the computations become much more difficult. Obtaining probability distributions and estimates of probabilities of hostile action from experts is discussed in Section 4.3. While maintaining current probability distributions agreed upon by experts may be the most difficult part of the problem, it should be feasible with the help of a user-friendly, interactive program.

6.0 REFERENCES

Carlson, D. R. September 1989. Letter under "Comment and Discussion" in response to N. Friedman, U. S. Naval Institute Proceedings, pp. 87–92. Reprinted as "It was avoidable, and the Navy must learn from it," The San Diego Union, September 10, 1989. [CMDR Carlson was Commanding Officer of the USS *Sides*, a frigate accompanying the USS *Vincennes*.]

Crowe, W. J., Jr. 10 September 1989. "Gulf experience, bad data led to cruiser's attack error," The San Diego Union, Reprinted from a letter of August 18, 1988. [ADM Crowe was Chairman of the Joint Chiefs of Staff at that time.]

Dempster, A.P. 1967. "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics*, no. 38, pp. 325–339.

Dillard, R. A. 23–27 August 1982. "The Dempster-Shafer Theory Applied to Tactical Data Fusion in an Inference System," *Proceedings of the 5th MIT/ONR Workshop on C^3 Systems*, Naval Postgraduate School, Monterey, CA, MIT-LIDS-R-1267, 1982, pp. 170–174.

Friedman, N. May 1989. "The *Vincennes* Incident," U. S. Naval Institute Proceedings, pp. 72–80.

Garvey, T. D., J. D. Lowrance, and M. Fischler, August 1981. "An Inference Technique for Integrating Knowledge from Disparate Sources," *Proceedings of the IJCAI 7*, vol. 1, pp. 319–325.

Genest, C. and J. V. Zidek. 1986. "Combining Probability Distributions: a Critique and an Annotated bibliography," *Statistical Science*, vol. 1, no. 1, pp. 114–148.

Lainiotis, D. G. and S. K. Park. 11–13 January 1972. "Feature Extraction Criteria: Comparison and Evaluation," *5th Hawaii International Conference on System Sciences,* University of Hawaii.

Meyer, M. and J. Booker. 1991. *Eliciting and Analyzing Expert Judgement: A Practical Guide*, Academic Press, London.

Ng, K. C. and B. Abramson. April 1990. "Uncertainty Management in Expert Systems," *IEEE Expert*, pp. 29–48.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Shafer, G. 1976. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.

Shortliffe, E. H. 1976. *Computer-Based Medical Consultations: MYCIN*, American Elsevier, New York, NY.

Tversky, A. and D. Kahneman. 27 September 1974. "Judgment under Uncertainty: Heuristics and Biases," *Science*, vol. 185, pp. 1124–1131.

Yager, R. R. 1977. "Multiple Objective Decision Making Using Fuzzy Sets," *International Journal of Man-Machine Studies*, vol. 9, pp. 375–382.

Zidek, J. V. 13–14 November 1986. "Group Decision Analysis and Its Application to Combining Opinions," *Workshop on Assessing Uncertainty*, Naval Postgraduate School, Monterey, CA, Stanford Univ., ONR Contract N00014-76-C-0475, 16 March 1987.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1991	Final: Oct 1990—Aug 1991

4. TITLE AND SUBTITLE

STATISTICAL DECISION MAKING WITH UNCERTAIN AND CONFLICTING DATA

5. FUNDING NUMBERS

PE: 0601152N
WU: DN300160

6. AUTHOR(S)

R. A. Dillard

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Ocean Systems Center
San Diego, CA 92152–5000

8. PERFORMING ORGANIZATION REPORT NUMBER

NOSC TR 1451

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Chief of Naval Research
Arlington, VA 22217–5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Automated decision-making techniques are investigated for use in situations where incoming tactical reports are frequently inaccurate or incorrect and where information about probability models is limited. The first problem addressed is to find a decision method for deciding among a set of hypotheses $\{Hi\}$ based on a set of evidence $\{Ej\}$, given data quality factors $\{Wj\}$ and numerical probability judgments such as $\{P(Hi|Ej)\}$ or $\{P(Ej|Hi)\}$. Four candidate decision methods are compared: the linear opinion pool, the logarithmic opinion pool, Dempster's rule, and a fuzzy-logic algorithm. Comparisons include the MYCIN certainty-factor calculus for the two-hypothesis examples. The decision methods give similar results, but none is fully satisfactory. At least one should perform adequately if selected carefully for the decision problem, but an effort should be made to find a fully suitable method. Several statistical techniques are investigated for measuring the relevance of evidence to a decision problem and for identifying suspicious or conflicting evidence. An approach to implementing the decision process in a knowledge-based system is outlined. The problem of eliciting and aggregating numerical probability judgments is also discussed.

14. SUBJECT TERMS

statistical decision making
data quality
uncertainty

15. NUMBER OF PAGES

54

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

UNCLASSIFIED

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE *(include Area Code)*	21c. OFFICE SYMBOL
R. A. Dillard	(619) 553-4140	Code 444

NSN 7540-01-280-5500

Standard form 298

UNCLASSIFIED

INITIAL DISTRIBUTION

Code 0012	Patent Counsel	(1)
Code 0144	R. November	(1)
Code 421	I. Goodman	(1)
Code 421	P. Calabrese	(1)
Code 421	F. Kramer	(1)
Code 423	K. Fernandes	(1)
Code 44	J. Grossman	(1)
Code 442	D. Bamber	(1)
Code 444	D. Eddington	(2)
Code 444	R. Dillard	(2)
Code 444	C. Haupt	(1)
Code 444	J. Herman	(1)
Code 444	R. Larsen	(2)
Code 444	K. Pickslay	(1)
Code 444	T. Wadsworth	(1)
Code 952B	J. Puleo	(1)
Code 961	Archive/Stock	(6)
Code 964B	Library	(3)

Defense Technical Information Center
Alexandria, VA 22304-6145 (4)

Center for Naval Analyses
Alexandria, VA 22302-0268

Naval Air Development Center
Warminster, PA 18974-5000

Naval Underwater Systems Center
Newport, RI 02841-5047

Navy Acquisition, Research &
Development Information Center
(NARDIC)
Alexandria, VA 22333

Fuzzy Logic, Inc.
La Jolla, CA 92037

NOSC Liaison Office
Arlington, VA 22217-5000

National Security Agency
Ft George G. Meade, MD 20755-6000

Naval Research Laboratory
Washington, DC 20375-5000

Naval Postgraduate School
Monterey, CA 93940

Navy Acquisition, Research &
Development Information Center
(NARDIC)
Pasadena, CA 91106-3955